Macroeconomic Analysis

Edward Shapiro

Fourth Edition

Macroeconomic Analysis

Fourth Edition

Macroeconomic Analysis

Fourth Edition

Edward Shapiro
UNIVERSITY OF TOLEDO

Harcourt Brace Jovanovich, Inc.
NEW YORK SAN DIEGO CHICAGO SAN FRANCISCO ATLANTA

To the memory of my father

Library of Congress Catalog Card Number: 77-89072

ISBN: 0–15–551212-9

Printed in the United States of America

Preface

The objective of this edition remains unchanged from that of its predecessors. However, in terms of the tools and techniques employed to achieve the objective, this edition differs significantly from its predecessors. These differences will be outlined below. As for the objective, it may be stated as in earlier editions: to provide a treatment of macroeconomic measurement, theory, and policy designed specifically for students whose background in economics is limited to the customary year of introductory survey work. Also, as stated in earlier editions, the treatment is of a kind that does not call on students to bridge the huge gap between introductory economics and graduate level macroeconomics in the one-quarter or one-semester course usually devoted to the subject. Nonetheless, this treatment will, I believe, enable students to secure a solid grasp of what are generally recognized as the fundamentals of macroeconomic analysis.

As in earlier editions, each topic is developed at sufficient length to make it meaningful to students. This reflects my belief that a brief outline type of presentation is of little lasting value to the typical student at this level and that it is worth taking the space and effort to spell out topics clearly and thoroughly. The favorable acceptance given to preceding editions suggests that many instructors have the same belief.

Also, as in preceding editions, the material is presented essentially in non-algebraic form. Some algebra is employed, but reliance on it in this edition is no greater and no less than in previous editions. It has been argued that there are no important ideas in economics that cannot be explained with carefully written English; written explanations and diagrams remain the basic tools employed in this book. However, a bow to the algebraic tool is found in the new edition of the student workbook that accompanies the text. An algebraic formulation of the text's *IS-LM* model with accompanying problems appears as an appendix to Chapter 16 of the workbook.

Among the major substantive changes introduced in this edition is a restructuring of the chapters on income determination in terms of aggregate demand and aggregate supply functions. With the hope that it will avoid the confusion that

results in some students' minds from the practice of attaching the same label to two completely dissimilar things, I have restricted the use of the term "aggregate demand function" to the relationship between the price level and the aggregate quantity of goods demanded and the term "aggregate supply function" to the relationship between the price level and the aggregate quantity of goods supplied. Apart from the application to output in the aggregate rather than in a single market, this usage has the advantage of paralleling the terminology the student meets in microeconomics.

In the simple Keynesian model developed in Chapters 4–7, the assumption is that the aggregate supply function is perfectly elastic over the range of output being produced so that the price level is stable. Because the output level on this assumption is determined by the position of the aggregate demand function, regardless of its price elasticity, the diagrammatic relationships between the aggregate quantity of goods demanded and the price level and between the aggregate quantity of goods supplied and the price level are ordinarily not shown at this stage of the analysis. However, in later analysis where the aggregate supply function is no longer assumed to be perfectly elastic over the relevant range, these relationships become the centerpiece of the analysis. I believe that students will better understand these particular relationships in the more advanced analysis of the later chapters if they see that such relationships also exist, although in a special form, in the simple analysis of the early chapters. Furthermore, showing these relationships in the simple analysis provides a constant diagrammatic reminder of the assumption that the price level is stable. This is especially important because this is the kind of assumption so easily forgotten by students whose only real-world experience has been with a price level that has been anything but stable.

The restructuring of the chapters on income determination in terms of aggregate demand and aggregate supply functions has required extensive rewriting of the material in Part Four. Chapter 13 is in large part new. It develops, in turn, the upward sloping, the perfectly inelastic, and the perfectly elastic (up to full employment output) aggregate supply curves. Chapter 14 on the basic classical model (Chapter 17 in the preceding edition) has been largely rewritten, and all the diagrams have been reconstructed to show explicitly the perfectly inelastic aggregate supply function that is part of this model. The *IS-LM* model that was covered in one chapter (Chapter 19) in the preceding edition is now the subject of Chapters 16–18 in Part Four. As in Chapters 4–7 that develop the simple Keynesian model or the model without money and interest, Chapter 16 assumes a fixed price level. The aggregate supply function is assumed to be perfectly elastic up to the full employment level of output, and the economy is assumed to be operating along that perfectly elastic portion. Chapter 17 drops the assumption of a perfectly elastic supply function and thus shifts the analysis to the case of a variable price level. Among other subjects, the conventional material on wage-price flexibility is found in this chapter. Chapters 16 and 17, like the *IS-LM* analysis in the preceding edition, are restricted to a closed economy. The completely new Chapter 18 extends the *IS-LM* analysis by introducing the balance of pay-

ments function (BP) and develops the *IS-LM-BP* analysis that is part of an open economy.

Also entirely new in this edition is the treatment of national income accounting in Chapter 2. This chapter is limited to the bare essentials of the subject with the detail made available in an appendix at the end of the book. Most of what was covered in the four-chapter treatment of national income accounting in the preceding edition has been revised and organized into this extensive appendix.

Apart from a few pages on the definition of inflation, Chapter 22 is also completely new. It covers in some detail the effects of inflation on the distribution of income, on the distribution of wealth, and on employment, output, and the growth rate.

Numerous other changes, ranging from a few lines to a few pages, are scattered throughout the book. Many of these are the result of comments provided by teachers who have been kind enough to write.

Some space for new material within the body of the text has been obtained by presenting most of the previous edition's coverage of national income accounting in an appendix. However, instructors may still find more in the body of the text than can be covered in a one-semester course or *a fortiori* in a one-quarter course. Because not all instructors have the same ideas about what is important, it seems to me better to have a somewhat longer book from which individual instructors may omit portions that they believe are less important rather than a shorter book that denies all who use it any choice of this kind. This being so, the following notes are offered as a guide to what chapters or major sections of chapters may be omitted without producing any significant breaks in continuity.

Probably all will want to cover Chapter 1, which provides some background on macroeconomic theory, although it is not indispensable. Similarly, Chapter 3 is not indispensable, but it is helpful to students who have a weak grasp of the basic concepts covered there.

The four chapters that make up Part Two are essential to what follows. However, if one wants to omit the foreign sector, which is sometimes covered in another course, one may omit Chapter 7. In this event, Chapter 18 must also be omitted.

Those who want to devote extra time to Part Four or to cover more of the material in Part Five may do so by omitting some of what is in Part Three. The concept of the consumption function in the minimum form needed for work in later chapters is covered in Chapter 4. Those who want to look at the record revealed by the times series on aggregate consumption and income but do not want to pursue the various consumption function hypotheses and their implications may omit all but pp. 125–34 of Chapter 8. Similarly, one may omit Chapter 9, which deals with the nonincome influences on consumption, without facing any problem in later chapters. Chapter 9 (apart from item 3, pp. 149–50, of the section on Distribution of Income) may be included even when Chapter 8 is not. Of the chapters on investment (Chapters 10 through 12), Chapter 10, which explains the marginal efficiency of capital and the derivation of the marginal efficiency of investment schedule, is probably not one that many would choose to omit. Furthermore, an understanding of much of the material on investment in Chapters

11 and 12 requires an understanding of the concepts and relationship between the marginal efficiency of capital and the marginal efficiency of investment. However, if one wants to reduce the amount of attention given to investment, the second part of Chapter 11 on the accelerator theory may be covered without the first part on the profits theory, and the second part of Chapter 12 on "The Role of Finance—Beyond the Interest Rate" may be covered without the first part of that chapter.

The chapters in Part Four must be studied in the order given. The first three chapters of this part provide a foundation for the last three, which develop what is called the extended model. This is the aggregate demand-aggregate supply model in which aggregate demand is derived from the *IS-LM* framework and aggregate supply is perfectly elastic up to the full employment output in the fixed price level case (Chapter 16) and either upward sloping or perfectly inelastic in the variable price level case (Chapters 17 and 18). As noted above, the last chapter in this part may be omitted in its entirety by those who want to limit attention to a closed economy. However, none of the other chapters in this part may be entirely omitted. Those who do not wish to work through the basic classical model in the detail provided by Chapter 14 may limit attention to pp. 232–36 without a serious break in continuity. These pages are needed as a minimum for an understanding of Chapter 15 in which the development of the Keynesian theory of the demand for money assumes an understanding of the simple quantity theory of money presented in these pages.

Considerable opportunity to "pick and choose" is available in Part Five. Chapter 19 includes the familiar multiplier-accelerator interaction, which can only be assigned if the accelerator theory section of Chapter 11 has also been studied. Also a prerequisite to the inclusion of the Kaldor model of the trade cycle in Chapter 19 are pp. 181–86 of Chapter 11. Within Chapter 19, one may cover the accelerator-multiplier interaction while omitting the balance of the chapter. Although Chapter 20 on the nature of growth provides a helpful background, it is not essential to the understanding of Chapter 21 on growth theory. In Chapter 21, the section on the Harrod-Domar theory which incorporates a theory of investment requires an understanding of the accelerator theory section of Chapter 11. On the subject of inflation, one may proceed from the first section of Chapter 22 on the definition of inflation to Chapter 23 on the process and causes of inflation. The major part of Chapter 22 on the effects of inflation is not needed to understand Chapter 23. If one chooses to cover Chapters 24 and 25 on policy, the only prerequisite in Part Five is Chapter 23 on the process and causes of inflation. Chapter 24 is devoted to incomes policy and fiscal policy and one may omit the incomes policy portion with no problem.

Although these notes will help one decide what can be omitted from the book as written, they do nothing for one who finds omitted from the book material that he or she believes should be included. Needless to say, anyone who uses this book can only find within its covers what the author has put there, and there is almost no limit to the choice an author has in this regard. I can only hope that what I have chosen provides a well-balanced coverage of what most teachers believe is appropriate for a course at this level.

I also hope that this edition of *Macroeconomic Analysis* contains fewer of the ambiguities and inconsistencies that seem to find their way into even the most carefully written accounts. For the correction of a number of these I am indebted to the instructors who have been kind enough to point out such difficulties in the preceding edition. I also wish to thank James R. Ostas who read all of the new material and much of the old and made many helpful suggestions. Neither he nor anyone else but me is responsible for the shortcomings that remain.

Edward Shapiro

Contents

Macroeconomic Analysis

Fourth Edition

part one

Introduction: Measurement and Concepts

chapter 1

The Background of Macroeconomic Theory

In an area that changes as fast as economics, anything that persists for more than a decade qualifies as "traditional." By this criterion, the division of the area of study called economic theory into two major branches, most commonly identified as macroeconomics and microeconomics, is now traditional, these terms having been in general use for several decades. Numerous subdivisions are found under these headings—monetary theory, consumer behavior theory, business cycle theory, production theory, wage theory, and the like—but the initial broad separation is into macroeconomic and microeconomic theory.

Several older terms are sometimes used to identify these two branches, but they have never become part of the language of economics in the way that macroeconomics and microeconomics have over the past decades. In fact, until the thirties there was little need to distinguish the two branches of economic theory, for economists concentrated their attention almost exclusively on what is now traditionally known as microeconomic theory. Macroeconomic theory was clearly the junior partner. However, a new interest in macroeconomics arose in 1936, the year of the publication of John Maynard Keynes's *The General Theory of Employment, Interest, and Money*,[1] the year of the beginning of a change so momentous that some choose to call it the "Keynesian revolution." With the ferment begun by the ideas in Keynes's book, economists' relative neglect of macroeconomic theory ended.

MACROECONOMICS AND MICROECONOMICS

"The term 'macro-economics' applies to the study of relations between broad economic aggregates."[2] "Macroeconomic theory is the theory of income, employment, prices and

[1] John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936. This work will hereafter be referred to as the *General Theory*.

[2] R. G. D. Allen, *Macro-Economic Theory*, St. Martins, 1967, p. 1.

money."[3] Macroeconomics is "that part of economics which studies the overall averages and aggregates of the system."[4] None of these statements, or any other short statement that could be given, satisfactorily defines the term, and each author just quoted follows his short statement with several more sentences, paragraphs, or even pages in an attempt to give the term a clear meaning. While they differ somewhat in emphasis, all such explanations bring out the idea that macroeconomics deals with the functioning of the economy *as a whole*, including how the economy's total output of goods and services and its total employment of resources are determined and what causes these totals to fluctuate. It seeks to explain why at some times as little as 3 percent of the labor force is unemployed and at other times as much as 9 percent or even more, and why at some times there is full utilization of the economy's productive capacity as measured by its workers, factories, equipment, and technological know-how and why at other times a good part of this capacity to produce goes to waste. It also seeks to explain why the total of goods and services produced grows at an average rate of 4 percent per year in one decade and at an average rate of 2 percent in another, and why in some time periods the price level rises sharply, while in others it remains stable or even falls. In short, macroeconomics attempts to answer the truly "big" questions of economic life—full employment or unemployment, capacity or undercapacity production, a satisfactory or unsatisfactory rate of growth, inflation or price-level stability.

In contrast, microeconomics is concerned, not with total output, total employment, or total spending for all goods and services combined, but with the output of particular goods and services by single firms or industries and with the spending on particular goods and services by single households or by households in single markets. The unit of study is the part rather than the whole. For example, microeconomics seeks to explain how the single firm decides the sale price for a particular product, what amount of output will maximize its profits, and how it determines the lowest cost combination of labor, materials, capital equipment, and other inputs needed to produce this output. It is also concerned with how the individual consumer determines the distribution of his total spending among the many products and services available to him so as to maximize his utility. In its approach, microeconomics takes essentially as *given* the total output, total employment, and total spending for all goods and services and proceeds to examine how output and employment are allocated among various individual industries and firms within industries and how the prices of the various products of these individual firms are established. Microeconomics asks how shifts in consumer spending from the product of one industry to that of another, or from the product of one firm within an industry to that of a competitor, will cause output and employment to be reallocated among different goods and services and among different industries and firms.

What microeconomics takes essentially as *given*—namely, the total output for the economy as a whole—is what macroeconomics takes as the prime *variable* whose size or value is to be determined. What macroeconomics takes as *given*—namely, the distribution of output, employment, and total spending among the particular goods and services of individual industries and firms—are all *variables* in microeconomics. In regard to prices, what microeconomics takes as given—namely, the general price level—macroeconomics takes as a variable; and what macroeconomics takes as given—namely, the relative prices or exchange ratios among individual goods and services—microeconomics takes as variables.

Although such a sharp distinction helps to clarify the essential differences, the preceding

[3] J. M. Culbertson, *Macroeconomic Theory and Stabilization Policy*, McGraw-Hill, 1968, p. 7.

[4] K. E. Boulding, *Economic Analysis*, Vol. II, 4th ed., Harper & Row, 1966, p. 1.

is admittedly a much sharper distinction than can legitimately be made. Strictly speaking, there is only one "economics." In practice, analysis of the economy is not conducted separately in two watertight compartments. As one analyzes macroeconomic variables and their relationships, one must also allow for changes in microeconomic variables that may have an impact on the macroeconomic variables and vice versa.[5] As one analyzes the economic processes that determine the nation's material well-being, one must consider both macroeconomic and microeconomic aspects. From the macroeconomic point of view, the nation's material well-being will be greater the closer the economy comes to *full utilization* of its total resources, taking as given the allocation, good to bad, of the amount of these resources that are actually employed in the production of the economy's output. From the microeconomic point of view, material well-being will be greater the closer the economy comes to *optimum allocation* of its resources, taking as given the degree of utilization, partial to full, of its total resources. Clearly, the basic goal is the same from both points of view: the maximum material well-being for the population as a whole. This goal can only be attained with both full utilization and optimum allocation of all available resources.

The distinction just drawn between macroeconomics and microeconomics helps account for the shift of emphasis from the traditional concern with microeconomics to concern with the economy as a whole, a shift of emphasis that is generally recognized as a Keynesian phenomenon. Before the thirties, economists emphasized microeconomics as if by default, for at the time it seemed that there remained little more to say about macroeconomics. The accepted macroeconomic theory of the day argued that total output was not really a variable but more in the nature of a constant in any short period, since its actual amount in any such period was simply whatever the fully employed economy could produce with the state of technology then existing. If this were indeed the case, the only relevant question would be whether these fully employed resources were being used to the best advantage—that is, whether they were optimally allocated among rival lines of production. However, if it were not in fact the case, the question of whether resources were being allocated to their best uses would lose much of its relevance. Optimal allocation of resources takes on the greatest importance when those resources are being fully utilized, for then there is truly a scarcity of resources. But in an economy operating substantially below full utilization, resources are not actually scarce, at least not at that time. To produce additional output of almost any kind under these conditions does not require that resources be diverted from the production of other kinds of output. The opportunity cost of producing additional output of almost any kind is close to zero as long as such output can be obtained by making use of otherwise idle resources. Therefore, to the extent that the economy departs from full utilization, macroeconomics takes on greater relative importance and microeconomics less relative importance.

Pre-Keynesian economic theory did not maintain that the economy remained uninterruptedly at full utilization or full employment. Its relative neglect of macroeconomics followed from its view that departures from full employment were strictly temporary. It argued that the automatic forces of competitive markets would carry employment and output back to the full-employment level in short order. As a consequence, departures from full employment did not generate widespread concern among economists as long as they believed that full employment was the normal situation to which the economy would automatically and promptly

[5] For example, to the extent that labor's geographical immobility is such that workers fail to move from an area whose industry is declining to another whose industry is growing, total output and total employment may be less than they would be with more mobility.

return. The fact that there were relatively few severe or prolonged depressions through at least the first half of the nineteenth century gave support to this belief. Lapses from full employment, being both infrequent and short-lived, could be easily explained away as exceptions to the full-employment rule.

Even though viewed as exceptions to the rule, the "panics," "crises," and "depressions" of the nineteenth century were nevertheless the subject of much study. David Ricardo wrote in 1817 of "revulsions in trade," and Mill in 1848 discussed "commercial crises" at length. However, the basic tenet underlying full employment as the norm went unchallenged by Ricardo, Mill and others. On the contrary, they staunchly defended it against the attacks of what few dissenters there were. This basic tenet was that aggregate demand for goods and services could not, other than temporarily, fall short of the aggregate supply of goods and services. As long as this was so, there was no barrier to the production of the aggregate supply that corresponded with full employment.

The handful of dissenters who attacked the rule that the economy automatically generated sufficient demand to absorb the supply produced under conditions of full employment failed, because none was able to buttress his attack with an alternative theory capable of supplanting the orthodox theory that aggregate demand could not be deficient. As James B. Conant observed, "It takes a new conceptual scheme to cause the abandonment of an old one."[6] Before Keynes's there had been no alternative theory that could be made to stick. Thomas R. Malthus, a dissenter of the early nineteenth century, attacked the accepted theory unsuccessfully, for, in Keynes's words, "Since Malthus was unable to explain clearly (apart from an appeal to the facts of common observation) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction."[7] Although there were a number of other attacks on the classical orthodoxy during these many years, a really successful one did not come until over a hundred years after Malthus in the form of Keynes's *General Theory*. Keynes presented an alternative theory of the determination of employment and output that explained why the forces of a market economy did not assure that aggregate demand would automatically be that which was necessary for full employment. He maintained that the level of employment was a variable, with full employment simply one possible level; an unlimited number of less-than-full-employment levels were also possible.

Keynes's *General Theory* offered an explanation of the economic disaster that the United States and many other countries suffered during the Great Depression, something the then existing body of theory was quite incapable of doing. The *General Theory* did not depend merely on "an appeal to the facts of common observation" to show that demand could be less than that needed for full employment; it supplied a theory to explain the facts, facts that were apparent to all in the early thirties. A "new" theory had come forth to deal with the reality of those depressed times.

During the decade following the appearance of the *General Theory*, economists addressed themselves to refining and building on the pioneer work of Keynes, to analyzing the complex economic processes that determine the actual level of employment—a level that the new theory showed could be one of full employment, one of severe unemployment, or any other level between these extremes. Keynesian theory was also applied during and after World War II to an analysis of inflation, a condition that was found to be closely connected with an economy at or near full employment. From such analysis came policy prescriptions designed to lift the system up to the full-employment level to which auto-

[6] James B. Conant, *On Understanding Science*, Yale Univ. Press, 1947, p. 89.

[7] *General Theory*, p. 32. See Keynes's Chapter 23 for his review of the "dissenters," or "heretics," of their time who are identified as the precursors of his own theory.

matic forces might fail to carry it and to maintain this level without inflation. For these reasons macroeconomics rose to prominence through the forties and fifties from its earlier relative obscurity.

By the mid-sixties the pendulum had begun to swing back toward the micro branch. It was not that there remained little more to say about macroeconomics, as the classical economists had believed a hundred years earlier, but there was a feeling that the major macroeconomic problems, even if not completely solved, were at least well under control. A series of earlier, less notable, policy successes had been capped with the impressive success of the 1964 tax cut; the large tax cut of that year, which had been urged by some economists as a means of getting the economy back to full employment and raising the lagging growth rate, produced very much the results that these economists had forecast. With the record before it, the profession was by then quite generally satisfied that the new economics of Keynes had given it the know-how to do a reasonably good job of getting the economy back to a full-employment growth path on the occasions when it slipped off that path. With basic macroeconomic problems believed to be in hand, a greater share of effort could be devoted to microeconomic problems. This was reinforced by the fact that problems like pollution control, energy supplies, consumer protection, crime control, and health care had rather suddenly taken on a much greater importance than before, and these problems are essentially microeconomic in their economic aspects.

The self-satisfaction that the economics profession felt in its macroeconomics branch was, however, short-lived. Although it may be that the boom and inflation that began in 1965 could have been prevented if the tax increases recommended by economists at that time had been put into effect, the process of curbing a boom and checking an inflation that had been permitted to gain momentum turned out to be more difficult and subject to greater uncertainty than had been thought to be the case by economists just a few years earlier. The increase in income tax rates finally put through in 1968 did not begin to have the dampening effect that economists had generally expected it to have. And beyond this, the various other restrictive policy actions taken over the next several years also were unable to check the inflation, although it turned out that they were able to push the economy into a recession in 1970.

Then, a few years later bad came to worse. The recovery from the 1970 recession had by 1973 reached boom proportions, but at the end of 1973, both in the United States and other major industrialized countries, began the sharpest recession experienced since the desperate days of the thirties. But the ultimate blow was not the severity of the recession itself; it was the fact that this grueling contraction was accompanied by one of the most rapid rates of inflation experienced since World War II. The term "double-digit inflation" became part of the language in 1974. There were, of course, special problems which are not amenable to domestic economic policy like the quadrupling of oil prices by the OPEC cartel in 1974, but the fact of sharply declining output and employment and sharply rising prices was the source of great discomfiture to the economics profession. Economists have the tools to deal with the malady of recession when that malady, as is usually the case, is not accompanied by an accelerating inflation. They also have the tools to deal with an accelerating inflation that occurs with the economy operating at full capacity. However, they do not now have the tools to simultaneously eliminate the twin evils of a sharp recession and a sharp inflation. In view of our experience over recent years, economists who during the sixties had believed that the tools were at hand to meet whatever macroeconomic problems might confront us in the future were convincingly disabused of such beliefs during the seventies. Solutions for the kind of macroeconomic problems which have faced us recently remain to be found; and macroeconom-

ics, which had risen from obscurity during the forties and fifties, is not about to return to obscurity during the late seventies for lack of basic problems with which to occupy itself.

MACROECONOMICS—CLASSICAL AND KEYNESIAN

Economic theory, as was noted, may be divided into macroeconomic and microeconomic theory; macroeconomic theory may in turn be divided into classical and Keynesian theory. But there is an all-important difference between these divisions, for whereas macro and micro theory are strictly *additive*, classical and Keynesian theory are largely *alternative*. One accepts macro and micro theory as parts of a total; one accepts parts of Keynesian theory only by rejecting parts of classical macroeconomic theory. One does not argue the relative merits of micro and macro theory, for they are largely separate, noncompetitive branches of theory, both of which are necessary to a complete theory of the economy. One does, however, argue the relative merits of classical and Keynesian theory, for their basic postulates and conclusions are opposed.[8]

What is meant today by classical macroeconomic theory? Marx, who coined the term "classical," used it to cover the theories of David Ricardo, James Mill, and their predecessors. Keynes extended the term to include "the *followers* of Ricardo, those, that is to say, who adopted and perfected the theory of the Ricardian economics, including (for example) J. S. Mill, Marshall, Edgeworth, and Prof. Pigou."[9] This is now the generally accepted meaning of "classical" in its application to macroeconomic theory. It is theory not attributable to any one person but distilled from the writings of many. What is more, the modern-day formulations of what is described as classical macroeconomic theory are nowhere to be found as such in the writings of the classical economists. These formulations had to be extracted from classical writings, since these writings did not specifically or systematically consider the basic questions raised by Keynes. Classical theory was not particularly concerned with the macroeconomic question of the level of employment because in classical theory automatic full employment was assumed.

What is meant by Keynesian macroeconomic theory? Here Keynes himself may be singled out as the founding father. However, his role is only this, since a distinction must be made between "the economics of John Maynard Keynes" and "Keynesian economics." The economics of Keynes, primarily his *General Theory*, is the foundation on which Keynesian economics has been constructed. Following the publication of his book, economists went

[8] Microeconomic theory has also undergone drastic change since the thirties, but this was change of a nature distinctly different from that which occurred in macroeconomic theory. Before the thirties, the theory of the individual firm, a core of microeconomic theory, assumed that the firm was either a perfect competitor at the one extreme or a monopolist at the other extreme. Either the firm was only one of so large a number of competing firms all producing an identical product that the individual firm had no appreciable control over the price charged for its product, or else the firm was the only producer of a product with virtually complete control over the price of its product. During the thirties the theory of monopolistic competition, or imperfect competition, was introduced to cover the wide range of market situations between perfect competition at the one extreme and monopoly at the other, market situations in which an individual firm had limited but appreciable control over the price of its product. The introduction of monopolistic, or imperfect, competition *added* something sorely needed to the body of microeconomic theory; it did not supplant or replace existing microeconomic theory. In contrast, the introduction of Keynesian macroeconomic theory was offered not to fill a gap in classical theory but rather to supplant or replace parts of that theory.

[9] *General Theory*, p. 3.

through it line by line, accepting, correcting, and rejecting. What they have built on the foundation that remained is a massive structure known as Keynesian economics.[10]

This serves only to distinguish classical theory and Keynesian theory on the basis of a few of the principal names. The important distinction, of course, has to do with the actual content of the theories. What in Keynesian theory is really novel and different, and what is merely an extension of classical ideas that were not spelled out by classical economists? What in classical theory was clearly wrong, at least when applied to the urbanized and industrialized economy of the twentieth-century world? These are extremely difficult questions that economists continue to debate. Still, what was advanced as "new" in the economics of Keynes and in the Keynesian economics that developed therefrom had an unprecedented impact on economic theory. The success of the Keynesian or "new" economics, as measured by its widespread acceptance, has few equals in the history of economic doctrine.[11]

This widespread acceptance of large parts of Keynesian economics over the past quarter-century has been equated with the widespread rejection of large parts of classical macroeconomics. Acceptance of a theory that showed that an economy may be in equilibrium with less than full employment amounted to a rejection of parts of classical theory that maintained that the only equilibrium position was the limiting one of full employment. This is the most fundamental of the differences in conclusions reached by the two theories.

The practical significance of this difference and what lies behind it cannot be exaggerated. For once we accept the Keynesian argument—that left on its own the economy may not move toward or attain the position of full employment—the way is open to the use of policies, designed to achieve this objective, that were ruled out by the quite different argument of classical theory. The ultimate purpose of any economic theory is to contribute to a sound basis for policy actions, and the widespread acceptance of Keynesian theory over the past two decades has increasingly led to the adoption of policies suggested by that theory. As many economists are fond of pointing out, the massive federal tax cut of 1964 amounted to an acceptance of Keynesian theory by the U.S. Congress. However, to obtain any real understanding of the rationale of this and other policy actions, one must first gain an understanding of the essentials of Keynesian theory.

The main task of this book is to develop the essentials of the Keynesian theory of output and employment, which has thoroughly dominated macroeconomic theorizing for almost a third of a century. This theory started out as a revolution against the classical orthodoxy and ended up with such success that by the fifties it had in turn become a new orthodoxy. In the normal course of events, revolution begets counterrevolution, and the Keynesian orthodoxy that evolved from the Keynesian revolution was itself in the sixties and seventies confronted with what has been called the monetarist counterrevolution.[12] It seems fair to say that some major parts of the pre-Keynesian orthodoxy have been permanently supplanted by the Keynesian revolution and the orthodoxy that evolved from its success, but there are

[10] In more recent years several economists have argued that this structure in certain fundamentals involves some misinterpreting of what Keynes really said. From this point of view, Keynesian economics differs from the economics of Keynes in part in its failure to correctly grasp the meaning of Keynes's words. This is a major theme of the book *On Keynesian Economics and the Economics of Keynes* (Oxford Univ. Press, 1968), by A. Leijonhufvud. See also R. W. Clower, "The Keynesian Counterrevolution: A Theoretical Appraisal," in F. H. Hahn and F. P. R. Brechling, eds., *The Theory of Interest Rates*, Macmillan, 1965.

[11] It is interesting that the success of Keynes was perhaps equaled by that of Ricardo, whose economics Keynes attacked. As Keynes put it, "Ricardo conquered England as completely as the Holy Inquisition conquered Spain." *General Theory*, p. 32.

[12] For an interesting essay on the economics and sociology involved in this development, see H. G. Johnson, "The Keynesian Revolution and the Monetarist Counter-Revolution," in *American Economic Review*, May 1971, pp. 1–14.

other portions of the pre-Keynesian orthodoxy that the monetarists have sought to revive, albeit in substantially revised form.

A cornerstone of the pre-Keynesian orthodoxy is the quantity theory of money, and the Keynesian success is in large part a measure of the defeat of that theory. A cornerstone of today's monetarist challenge is what is called the "modern" or "new" quantity theory of money. Although it differs from the old theory to such a degree that some argue that it is closer to Keynesian theory than to the old quantity theory, it does maintain the old theory's critical emphasis on the quantity of money in the economy as the supreme determinant of the economy's aggregate spending and, through this, of the economy's income, output, and employment levels. This sort of emphasis is violently anti-Keynesian—Keynes argued, and most economists eventually came to agree, that the tight causal linkage between changes in the quantity of money and changes in total spending that was indicated by the old quantity theory did not exist. To Keynes and to most economists since Keynes, the explanation for the changes in total spending involves a number of elements, and the quantity of money is only one, and not the most important one, of these elements. Most of this book will be devoted to the explanation of these elements and the theory built out of them by Keynes and refined and elaborated on by his followers.

At various points in the book we will touch on the modern quantity theory of money and on other aspects of monetarism. But the main focus remains on the Keynesian theory, which, so far at least, enjoys the support of a large majority of professional economists. What lies in the future is another matter. In the words of Professor Harry Johnson, "The question of interest is whether the monetarist counterrevolution will sweep the board and become the orthodoxy of the future, itself ripe for attack by a new revolution, or whether it will gradually peter out."[13] His personal answer is that he expects it to peter out. In this he probably expresses the expectation of the majority. The dedicated antimonetarists among this majority do not miss a chance to attack. Following a recent period in which total spending departed radically from what was expected by the monetarists, a *Business Week* story appeared under the heading: "The Critics Ask: Is Monetarism Dead?"[14] However, the monetarists have done their share of attacking and also have enjoyed a share of victories. The question posed by Professor Johnson at the beginning of the seventies had by no means been resolved by the late seventies.

It seems safe to say that anyone who looks into the background of macroeconomics today and for at least the next few years will conclude that macroeconomics may still be divided historically into two major periods: the pre-Keynesian period of the classical orthodoxy and the period of the Keynesian orthodoxy that was firmly established by the fifties. He will observe that the Keynesian orthodoxy faced a challenge that began in the sixties and in the debate that followed was forced to recognize and correct its intellectual shortcomings, but he will not conclude that the Keynesian theory was dislodged by the challenge of monetarism in anything resembling the way that the classical theory had been dislodged by the revolution begun by Keynes. It will, in other words, be granted that the monetarist attack led to some substantial qualifications and amendments to Keynesian theory but no more than that. It has been true for several decades, and remains so today, that an introduction to the macroeconomics branch of economic theory is essentially an introduction to Keynesian theory.

[13] *Ibid.*, p. 12.
[14] *Business Week*, June 7, 1976, pp. 63–64.

chapter 2

National Income Accounting

Ordinary business accounting seeks to summarize a firm's performance by measuring its profit or loss over a specified period of time. Similarly national income accounting seeks to summarize a country's economic performance by measuring its aggregate income and output of goods and services over a specified period of time. Although national income accounting later came to measure much more than this, in the early thirties it did little more than provide figures for the economy's aggregate income. It was the new interest in macroeconomic theory sparked by the publication of Keynes's *General Theory* in 1936 that was largely responsible for the development of national income accounting from its rudimentary form in the early thirties to the advanced form it had reached in the decade following World War II. This development included a more elaborate framework of accounts and initiated the collection of data for many income and output measures that appeared for the first time in this expanded framework.

Long before the nineteen-thirties, economists and others had been interested in knowing each year's aggregate income and output as well as its composition, but that interest was probably less than it came to be in later years. As long as the economy appeared to show no more than short-lived, moderate departures from full employment, detailed figures for its actual income and output from period to period were less important than they would be for an economy that showed greater instability. If there are indeed long-lived and severe departures from full employment, as during 1929–33, such figures are the means by which one can tell how bad things have gotten. It is interesting that the first national income estimates prepared by a U.S. government agency were produced in the thirties largely for this very reason.

Not only did the advent of the Keynesian theory speed the development of national income accounting but it also influenced the form of that development. The conceptual framework in which the accounts are presented was deliberately designed so as to facilitate the study of macroeconomic problems undertaken with the set of tools provided by Keynesian theory. For example, beyond simply showing the total output of the economy, one may break that total down in various ways, e.g., by industry of origin, by kind of output (durable goods, nondurable goods, services, and structures), and in other ways. However, because Keynesian theory made aggregate demand the key to explaining aggregate output and approached the explanation of aggregate demand by considering the spending of consumers, businesses, government, and foreigners, the classification

of the economy's aggregate output into the parts obtained by each of these four groups is the classification one finds at the core of the accounting framework.

This accounting framework in its entirety offers a systematic picture of the economic structure and process in terms of the interrelated flows of income and product, which are the basic variables of the economic process itself. Although the accounting data supplied quarterly for the interrelated income and product flows do nothing more than identify the changes that have occurred, a study of these changes enables economists to explain why they may have occurred and, more importantly, to forecast changes that may occur in the future. For example, given only the quarterly figures for the spending flows of the four groups, namely, personal consumption expenditures (*C*), private domestic investment (*I*), which is broken down into fixed investment and the change in inventories, government purchases of goods and services (*G*), and net foreign purchases, equal to exports (*X*) minus imports (*M*) or ($X - M$), a series of quarterly increases in inventories whose size appears to be far out of line with the increases in other flows such as consumption expenditures suggests the possibility of a recession ahead to correct what seems to be excessive inventory accumulation. The economist could reach a tentative judgment like this by looking only at the data provided in one part of the overall accounting framework, the National Income and Product Account, which shows the breakdown of aggregate output into the kinds noted above. More involved examples could be given, but these would require introducing other parts of the overall accounting framework.

For the macroeconomic theory in the following parts of this book, it is not essential for us to examine that framework in its entirety. The complete framework is composed of a set of interconnected accounts—a Personal Income and Outlay Account which shows the various sources of income of persons and the disposition of that income, a Government Receipts and Expenditures Account which does the same for all governmental units, a Foreign Transactions Account which shows receipts from foreigners for sale of goods and services to them and payments to foreigners for purchases of goods and services from them, implicitly a consolidated Business Income and Product Account which shows output produced and income generated within all business firms consolidated, and a Gross Saving and Investment Account which lists the amounts of gross saving of different sectors on one side and the matching amount of gross investment on the other side. Finally, the National Income and Product Account, brings together information from the other accounts and shows on the product side the various output flows that add up to the economy's final product and on the income side the various flows that add up to the same total found on the product side. The National Income and Product Account is a kind of summary account for the over-all framework. This chapter focuses on that account and on the principles that lie behind the measurement of the basic income and product flows found in it. The reader who wants to go beyond this to study the complete framework and the separate sector accounts out of which it is built, as well as the details behind the accounts, is referred to the long appendix at the end of the book which builds on the foundation provided by this introductory chapter.

INCOME AND PRODUCT

One fundamental identity in conventional business accounting is the balance sheet identity: Assets = Liabilities + Net Worth. In presenting the balance sheet, assets are listed on one side and liabilities and the various capital accounts that make up net worth are listed on the other

side. An equally fundamental identity in national income accounting is that between "income" and "product." The National Income and Product Account lists on one side the various income flows whose total is equal to the total on the other side, which lists the various product flows. Just as the two sides of the balance sheet show the same total because the terms are defined so as to assure that outcome, the two sides of the National Income and Product Account give the same total for the same reason. There is a dollar of income generated for every dollar of goods and services produced in the economy or a dollar of goods and services produced for every dollar of income generated in the economy. Just as it is customary in the accounting framework to refer to goods and services as product, we may also say that there is a dollar of income for every dollar of product or a dollar of product for every dollar of income.[1]

This means that once we have adopted a consistent set of definitions of income and product, we can arrive at an estimate of the value of the economy's output for any time period from either side: by adding up the value of all of the goods and services which are counted as product or by adding up all of the income generated in the course of producing all of these goods and services. Because the information on each side is wanted not only for the total it reveals but for other purposes, both income and product estimates are made in practice.[2] The estimates on the product side are based essentially on the expenditures made for product, so the derivation of estimates on that side is usually called the expenditures approach. The estimates on the income side are estimates of income itself and the derivation of estimates on this side is usually called the income approach.

These two approaches to the measurement of the value of the economy's output, as well as some of the conceptual problems faced in measurement, are examined in turn in the following pages. There is more than one major measure of the value of the economy's output in the U.S. accounting system—actually there are three—and there are also some other measures that will be essential in our later work, even though they are not measures of the value of output. Each of these will be identified and the relationships among them will be traced. A final section of this chapter lays out the basic national income accounting identities, which are also essential for our later work.

Measuring the Value of the Economy's Output of Goods and Services: The Expenditures Approach

One possible way to arrive at a figure for the value of the economy's aggregate product for a given time period would be to get figures on the number of units of each of millions of different goods and services produced during that time period (each different size and quality of a "good" is in effect a different good), gather up thousands of price lists, and proceed with the necessary multiplication of quantities by prices and add these to obtain a total. Even assuming all these data were available, it would be virtually impossible to work a mass of data in this form into a summary figure which could be described with any accuracy as the value of the economy's output for the time period. However, a summary figure of this kind can be reached in a much more direct way and with far more accuracy by counting each unit of each good and service produced at the time it is

[1] As will be seen below, the identity between income and product is based on a broad definition of income which includes flows like indirect taxes and depreciation charges not usually thought of as income. However, this broad definition of income is most convenient at the outset. Once the structure of the National Income and Product Account is grasped, one has no problem in switching to a narrower definition of income for which the identity is no longer income equals product but income plus certain nonincome flows equal product.

[2] It was in part because of the limitations imposed by the available data that the first governmental estimates of the value of the economy's aggregate output in the early nineteen thirties were derived on the income side only. During World War II these estimates were derived for the first time on the product side as well.

purchased and valuing it at the actual purchase price. Thus, if we add up total expenditures on goods and services, we will, subject to a number of qualifications to which we will turn next, have the total we seek, and also a total in which each unit of output is valued at what appears to be the best available indicator of the value of that unit to society, namely, the price actually paid for it by its buyer.

Exclusions from Expenditures This general approach to measuring the value of output through expenditures is perfectly proper and manageable, but it must obviously be refined into something more than simply getting a total which includes expenditures for all goods and services without qualification. Because the total being sought is a figure that measures the amount of output produced by the economy during a specified time period, it must clearly include only expenditures on the purchase of goods and services produced during that time period. It must not include any part of the many billions of dollars spent during that time period on goods produced in *earlier* time periods. All expenditures of this kind reflect no more than changes in the ownership of preexisting output and as such are not part of the total of expenditures that measures the value of current output.

Also excluded must be all expenditures for anything which is neither a good nor a service and which therefore does not reflect production at all, either current or past. For example, people in each time period spend billions of dollars on stock and bonds, some of which are purchased from other people who acquired them in earlier years and some of which are newly issued by corporations during that time period. Whether currently issued or not, these billions of dollars must *not* be counted as so many billions of dollars of expenditure in the total that is being built. There is no production or output of goods and services corresponding to expenditures for mere pieces of paper.

Somewhat less obviously but for the same basic reason, one must also exclude all expenditures by federal, state, and local government *for which the government does not receive a good or service in exchange.* The huge amount of government expenditures for things like social security, unemployment compensation, and veterans' benefits is not spending for a good or service; the recipients of these amounts spent by government do not provide a good or service to the government in exchange. This type of expenditure, so-called government transfer payments, must therefore be excluded, again for the same reason: it does not reflect output of goods and services and therefore its inclusion would rob the total expenditure figure which is being derived of its usefulness as a measure of the value of the economy's output of goods and services.

Intermediate vs. Final Product Even after excluding expenditures of the kinds indicated, we still cannot include all of the remaining dollar amounts spent on the purchase of every last kind of good or service produced during the time period. Such a total would be one that counts some goods or services in the total output once, others twice, still others three times, and so forth; this total would therefore drastically overstate the economy's true total output, whatever that happened to be.

The objective is to count, for example, the dollar value of the amount of bread that was produced during the period but not to count the dollar value of the amount of flour that went into the output of that bread, even though that amount of flour was produced by millers and sold to the bakers during the same time period. By the same reasoning, we do not want to add to the dollar value of the amount of bread the dollar value of the amount of wheat that went into the output of the flour that went into the output of the bread, even though that amount of wheat was produced by farmers and sold to millers during the same time period. At this point the wheat in the bread would be counted three times. Counting expenditures for the bread

alone counts the wheat only once, and this is what is wanted. Thus, in order to include in the total for the time period only that output which may be viewed as so-called *final product*, e.g., bread, all other output must be excluded which, although produced during the time period in question, is output, e.g., wheat and then flour, used up in the course of producing that time period's final product. Such excluded output is described as *intermediate product* in contrast to the included goods and services described as final product.

Distinguishing Intermediate and Final Product

The rationale for excluding intermediate product is clear enough in principle but extremely difficult in application. Unlike spending in the simple wheat–flour–bread case, there are large amounts of spending that do not fall distinctly into one or the other of the two categories, final product and intermediate product. Many government purchases of services fall into the gray area. For example, should the billions of dollars spent each year by state and local governments to provide the public with police protection be classified as spending for a final product and therefore counted as part of the economy's output or should it be viewed as spending for an intermediate product, that is, as spending for something "used up" in the course of producing a final product? If we view police protection narrowly as merely something without which the private sector of the economy would be unable effectively to operate its factories, offices, and stores, then police protection takes on the characteristic of an intermediate product, of something used up in the course of producing the output that comes out of the factories, offices, and stores. On the other hand, if we think of it as a service that exists apart from the goods and services produced by the factories, offices, and stores, then it is more in the nature of a final product.

It should be noted that the issue here is distinct from the earlier issue of government spending for goods and services versus government spending in the form of transfer payments. Here there is no question but that the recipients of the government spending, e.g., police officers, render a service in exchange—there is production of a kind corresponding to the government spending. The question is whether this service or thing produced is more in the nature of an intermediate product like flour that is used up in making bread or a final product like bread itself.

Questionable cases between final and intermediate product are by no means limited to government spending. For example, expenditures by consumers for transportation to and from work, for medical care, and for many other purposes sometimes described as "regrettable necessities" may perhaps be more appropriately classified as spending for intermediate product than for final product. A person's spending to get to and from his or her place of employment may be viewed as spending for a service that is "used up" in the course of producing the good or service that comes out of the place of employment and to whose production he or she contributes. Similarly, spending for medical care may be viewed as spending necessary to the maintenance of health and therefore necessary to the performance of a job at the place of employment and, as such, it is something "used up" in the course of producing the good or service that comes out of the place of employment.

There is an almost endless number of such cases. It might seem at first that the best approach would be to classify them one by one as cases of intermediate product or final product through a careful evaluation of the "merits" of each. However, the evaluation of each, no matter how carefully done, is still largely subjective and equally competent persons will not reach the same conclusion in each case. Although it is quite arbitrary, an alternative that avoids this problem of subjective evaluation is to treat all spending by both governmental units and individuals as spending on final product, as long, of course, as it is spending for

which the governmental unit or the individual receives a currently produced good or service in exchange. Despite the fact that this alternative is less than ideal, it is the approach taken in preparing the official governmental estimates of the economy's output in the U.S. and in some other countries. In these estimates, personal consumption expenditures (C) and government purchases of goods and services (G) each include all expenditures on goods and services by individuals and governmental units, respectively, subject to the qualifications noted.

Final Product: C + I + G To reach a total for spending on final product by consumers, businesses, and governmental units, i.e., a total for $C + I + G$, we must obtain a figure for the amount of business spending on goods and services that is to be included. Of all the spending by business for goods and services during a given time period, how much is to be excluded as intermediate product or, what is the same thing, how much is to be included as final product? To work toward an answer, let us go back to the simple illustration of the production of wheat, flour, and bread. The production of flour requires more than just wheat—among other things, millers need buildings and within them specialized kinds of machinery. Over a period of time, millers as a group will, of course, be buying wheat, electric power, pens and pencils, insurance coverage, and other nondurable items. In any time period, some millers will also be buying newly constructed buildings and/or new milling machinery. Now, over any period of time not only does each miller use up wheat, electric power, and the like in producing flour, but he also uses up plant and equipment as these durable assets gradually wear out with use over time. To find out then how much of all the goods and services purchased by our millers as a group during the period is *not* used up during the period in question, it would appear that all one has to do is find out how much they spent for new plant and equipment during that period and subtract from this the amount of plant and equipment they "used up" during the course of the period's production. An approximate figure for the amount of this "wear and tear" on plant and equipment for any period is provided by the amount that the firms charged in their books as depreciation expense for the period. The gross amount spent for new plant and equipment is called *gross fixed investment* and the net amount arrived at by subtracting the figure for plant and equipment used up is called *net fixed investment* of the millers for the period.

What is done for the millers is done for every other industry: gross and net fixed investment spending are determined for each and the totals for all industries combined are the gross fixed and net fixed investment for the economy as a whole. These may be designated as I_{gf} and I_{nf} in which *gf* and *nf*, respectively, mean gross fixed and net fixed. The figure for I_{nf} then gives us what appears to be a measure of the amount of total business spending for final product or the amount of total business spending that is not for goods and services used up in producing other goods and services during the time period. If we now add to the figures for spending by persons, C, and spending by governmental units, G, the figure found for I_{nf}, we will presumably have a total for spending on final product.

However, while this total is a measure of spending on final product by the three sectors, it is a total that from one period to the next will either overstate or understate the actual final product accounted for by the three sectors. As this total is obtained by counting all of the spending for final goods and services during the period, it will equal the actual total being sought only if there is an exact correspondence between the amount of final goods and services produced and the sum of C, G, and I_{nf}. The problem is that the total of these three types of expenditure need not and, as a practical matter, will not be the same as the amount of final product actually produced, because there may be, for example, some goods produced during the period that remain unsold during the period.

which means that there is no spending or final sale to reflect their production. This excess of what was produced over what was sold during the period will appear as an increase in business inventories, and an estimate of this amount must be added to the total derived for expenditures to reach the figure for the economy's output of final goods and services.

In the above example, it cannot be assumed that all of the millers' spending for goods and services, other than the amount measured as their I_{nf}, is for goods and services used up in the course of producing flour during the period. If at the end of the period they have a larger inventory of wheat on hand than at the beginning of the period, the value of this increase must be counted as part of final product for the period. If it had not gone into inventory but instead into the making of flour and then into the making of bread purchased by consumers, it would have been counted under C. However, it has actually gone into inventories and must be counted in that form or it will escape counting altogether and cause that much of an understatement of final product. By the same argument, if the millers show an increase in the inventory of their end product, flour, this increase too must be counted as part of final product. Or, in general, if the millers as a group show an increase in the value of their combined inventories of raw materials (wheat), goods in process, and end product (flour), this amount must be counted as part of final product. This incidentally, brings out the frequently misunderstood fact that *final* product for any time period is not identical with *finished* product for that period—an increase in business inventories of raw materials and goods in process must be included as part of the period's final product, although such goods plainly don't qualify as finished product by the usual meaning of that word.

Adding the change in the millers' inventories to their net fixed investment yields the correct figure for the amount of the economy's final product accounted for by investment within this industry. The figure thus obtained is called net investment (with the qualifier "fixed" dropped) I_n, and measures the sum of the amount of goods that millers purchased from other firms (like wheat and flour-making machines) that was not used up in the course of producing other goods during the time period and the amount of goods produced (like flour) that were not sold to other firms during the time period. An addition to inventories is thus like an addition to plant and equipment—each represents something produced during the period, not used up in further production during the period, and consequently something found in the possession of business at the end of the period. Accordingly, the two may logically be combined to form a total called net investment. If the change in inventories is combined with gross fixed investment, the total is, by the same reasoning, called gross investment (with the qualifier "fixed" dropped), I_g.

The change in the millers' inventories for a time period may, of course, also be negative. In this event, to arrive at the figure for I_n, the change in the millers' inventories must be subtracted from I_{nf}. If it also happened that the amount of plant and equipment used up by the millers during the period exceeded the amount of their spending for new plant and equipment, I_{nf} would be negative as well. The amount of the economy's final product accounted for by net investment within this industry or the sum of net fixed investment and the change in inventories within this industry would then be negative. The amount of gross investment accounted for by this industry would be the figure obtained by combining the change in inventories and gross fixed investment. The latter figure is the millers' total spending for plant and equipment and, unlike net fixed investment, cannot be less than zero. Although a negative figure for gross investment in a given time period can result from a decrease in an industry's inventories larger than its gross fixed investment spending, this would be a rarity for any single industry and does not occur at all for all industries combined.

What is done for the milling industry must be done for all industries. If for any particular time period all industries combined show an increase in inventories, the economy's final product will exceed the total of consumption spending, government spending, and business net fixed investment spending by that amount, which must accordingly be added to the total of the spending flows to obtain the total for final product. Similarly, if there is a decrease in inventories, that has to be subtracted from the total of the spending flows. As the total of the spending flows is commonly called final sales, the total for final product is equal to final sales plus the change in business inventories.

Final Product: Net Exports Added The sum of $C + I_n + G$ would be the final product for the economy as a whole only if the economy were closed. However, every economy actually imports some goods and services from the rest-of-the-world and exports some goods and services to it. If during a time period the domestic economy on balance exports more than it imports, this difference is part of the domestic economy's final product for the period. However, this part will not be picked up in the course of deriving the total spending of the three sectors within the economy and in working out the estimate for the change in inventories. This amount will be included only by adding it to the spending of the three domestic sectors.

For a numerical illustration, suppose that spending on final product by the three domestic sectors totals \$98 (all figures in billions), the change in inventories +\$2, the economy's exports of goods and services \$6, and imports \$5. Included in the \$100 billion sum for spending by the three domestic sectors and the change in inventories will be \$5 spent by these three sectors on imported goods and services, so that only \$95 of this total of \$100 is matched by domestically produced goods and services. However, \$6 of domestically produced goods and services have been exported and this amount must be added to the \$95 of domestically produced goods and services that remained within the economy to get a total for the domestic economy's output. The total is accordingly \$101. It may also be seen to be equal to the sum of the total spending of \$98 by the three domestic sectors, the +\$2 change in inventories, and the net export balance of \$1 or exports of \$6 minus imports of \$5.

For the opposite case in which the domestic economy imports more goods and services than it exports, that difference is a part of the output of other countries. However, it is output that will be counted as part of the domestic economy's output, if all we do is derive the total for the spending of the three sectors within the economy and add to this the change in inventories. To avoid this overstatement, the amount in question must be subtracted from the sum of the total spending of the three domestic sectors plus the change in inventories.

To illustrate, assume the same figures of \$98 for spending by the three domestic sectors and +\$2 for the change in inventories, but assume now that imports were \$6 and exports \$5. We then find in the \$100 total (\$98 + \$2), \$6 of imported goods and services, which means that only \$94 out of the \$100 reflects domestically produced goods and services. To this we must add \$5 of domestically produced goods and services that were exported to give us a total for the domestic economy's output of \$99. The total is again seen to be equal to the sum of the total spending of \$98 by the three domestic sectors, the +\$2 change in inventories, and the net export balance of −\$1 or exports of \$5 minus imports of \$6.

To sum up, inclusion of the economy's net export or import balance allows for the fact that total spending on final goods and services by the domestic sectors will not correspond with the total of the economy's final sales of goods and services. To estimate the value of the economy's output of final goods and services by means of the expenditures approach, one must add up the spending on final product not only by the three domestic sectors into which every domestic purchaser is classified

but also the net spending, positive or negative, on the domestic economy's product by the foreign sector. The sum of these four spending flows for the time period plus the change in inventories for the period will be the total sought.

Gross and Net National Product; Gross and Net Domestic Product According to Department of Commerce estimates, the sum of the expenditure or spending flows plus the change in inventories, known as gross national product, GNP, was \$1,706.5 billion for 1976. As may be found on the right side of Table 2-1, the four expenditure flows, C, I_{gf}, G, and $(X - M)$, were, respectively, \$1,094.0, \$230.0, \$361.4, and \$7.8 billion, which gives a total for final sales of \$1,693.2 billion. Adding to this the change in inventories of \$13.3 billion yields the figure of \$1,706.5 billion for GNP. As gross investment, I_g, is the sum of gross fixed investment and the change in inventories, GNP may be shown as the sum of the four flows, $C + I_g + G + (X - M)$ in which I_g is \$243.3 billion. This conforms with the breakdown of the GNP aggregate in Table 2-1.

The GNP aggregate is the most familiar concept, but it is one that includes some recognized "double counting" by failing to deduct from total business spending on capital goods, i.e., on structures and durable equipment, an

TABLE 2-1
National income and product account, 1976
(billions of dollars)

	Compensation of employees			1,036.3
	Wages and salaries		891.8	
	Supplements		144.5	
	Proprietors' income			88.0
	Rental income of persons			23.3
	Corporate profits			128.1
	Profits before tax		156.9	
	Profits tax liability	64.7		
	Dividends	35.8		
	Undistributed profits	56.4		
	Inventory valuation adjustment		−14.1	
	Capital consumption adjustment		−14.7	
	Net interest			88.4
NI	National Income			1,364.1
	Business transfer payments			8.1
	Indirect business tax			150.5
	Less: Subsidies less current surplus of government enterprises			0.8
	Statistical discrepancy			5.5
NNP	Charges against Net National Product			1,527.4
	Capital consumption allowances with capital consumption adjustment			179.0
GNP	Charges against Gross National Product			1,706.5

C	Personal consumption expenditures		1,094.0
I_g	Gross private domestic investment		243.3
	I_{gf} Gross fixed investment	230.0	
	Change in business inventories	13.3	
G	Government purchases of goods and services		361.4
$X - M$	Net exports of goods and services		7.8
GNP	Gross National Product		1,706.5

SOURCE: *Survey of Current Business*, U.S. Department of Commerce, July 1977.

Return to Capital.

allowance for the amount of such goods that were used up in the course of producing the year's total output of goods and services. The total of business spending on capital goods is the I_{gf} figure of \$230.0 billion and the amount of the existing stock of such goods used up during the year is \$179.0 billion as shown by the last entry on the income side of Table 2-1. If \$179.0 billion is subtracted from \$230.0 billion, the remainder is I_{nf} of \$51.0 billion or the amount of capital goods added to the preexisting stock during the year. Just as subtracting the figure for capital goods used up from gross fixed investment converts the latter figure into net fixed investment, it converts the total at the bottom of the right side from GNP to NNP. NNP is accordingly \$1,706.5 billion minus \$179.0 billion or \$1,527.4 billion (adjusted for rounding). The only difference between the GNP and NNP aggregates is the amount of capital consumption allowances.

As will be explained more fully in the next section, since by definition the product side and the income side of the account are identical, a broader definition of product like GNP is accompanied by a broader definition of income, and a narrower definition of product like the NNP is accompanied by a narrower definition of income. For every dollar of product there is a dollar of income: the exclusion from product of the amount of capital goods used up during the period is accompanied by the exclusion from income of the amount otherwise set aside to replace the capital goods consumed. Otherwise expressed, if product is defined to exclude expenditure for the replacement of capital goods used up, it is appropriate to define income to exclude an amount equal to the amount that would be generated in the production of the excluded amount of capital goods.

In the preceding pages the focus has been on I_{nf}, fixed investment net of replacement which is a nonduplicative figure, and correspondingly the focus has been on NNP, also a nonduplicative figure. However, over the years the GNP figure of Table 2-1 has been the one ordinarily presented to the public and used by economists as the summary figure for the economy's "final" output, despite the duplication or double-counting it contains. The statisticians derive the NNP figure in the way above noted: by subtracting the figure for capital consumption allowances from the figure for GNP. This means that the NNP figure from one year to the next will reflect any inaccuracy in the estimate of the amount of capital goods used up, but the GNP figure will not be affected at all by that error. The technical difficulties faced in estimating the amount of actual wear and tear or capital consumption have been such as to make this estimate one of the least reliable in the accounts. Therefore, to avoid the errors that would otherwise be introduced, the Department of Commerce has long emphasized the GNP rather than the NNP aggregate and correspondingly the I_{gf} rather than the I_{nf} figure. However, continuing work on the problem at long last led to the publication in 1976 of new estimates of capital consumption that are viewed as sufficiently accurate to permit meaningful measures of I_{nf} and NNP.[3] Therefore, in the years ahead the concept of NNP may gradually come to be as widely recognized by the general public as the concept of GNP has long been.

Both GNP and NNP include the value of the goods and services produced by labor and property supplied by residents of the U.S., whether that labor and property is located in the U.S. or in other countries. For certain purposes, it is useful to have measures that exclude the value of goods and services produced by the portion of this labor and property located abroad. If one deducts from GNP the net inflow of income earned on labor and property supplied by U.S. residents abroad (i.e., the gross amount so earned minus the amount earned by foreign residents on their labor and property located in this country), the remainder is called gross domestic product or GDP.

[3]The new measures appeared as part of a periodic comprehensive revision of the national income and product accounts published in the *Survey of Current Business*, January 1976.

As will be seen in the following section, the term *national income* is the one used in the U.S. accounts for the total of incomes earned by labor and various kinds of property known as factors of production, and the amount deducted from GNP to obtain GDP is the portion of the economy's national income that originates in the rest-of-the-world. In 1976 that was $14.4 billion, so that GDP was $1,692.1 billion or $1,706.5 billion minus $14.4 billion.

If one starts with NNP instead of GNP, a figure for net domestic product, NDP, is obtained in the same way. In 1976 NNP was $1,527.4; therefore NDP was $1,527.4 minus $14.4 or $1,513.0 billion. So far GNP is the more widely used of the two national measures, GNP and NNP. Similarly, GDP is the more widely used of the two domestic measures, GDP and NDP.

Measuring the Value of the Economy's Output of Goods and Services: The Income Approach

Because each dollar's worth of goods produced is matched by a dollar of income, we could arrive at the same figure for the value of the economy's output of goods and services on the income side that we reach on the product side, assuming there are no errors, deficiencies, or inconsistencies in the data or in the estimating procedures. In this section we want to look at the principal steps that must be taken and the principal problems that must be resolved in estimating the value of the economy's output from the income side. Much of what was said in the preceding part of this chapter applies here with appropriate modifications. Just as the approach on the expenditures side requires careful specification of what is to be included under the heading of expenditures, so the approach on the income side requires the same for what is to be included under the heading of income. In general, what is included must clearly be limited to those particular income flows which originate with the production of the goods and services whose total we seek to estimate, for only then will the total of the income flows be equal in value to that total of goods and services.

Exclusions from Income Specifically, this means that we cannot count on the income side the billions of dollars received by persons who sell buildings, automobiles, or any other good produced in an earlier time period, because what they receive in payment is not "income" in the sense of something generated in the course of producing the output of the current period. Actually, because such transactions involve a receipt of money from a sale of *assets*, such receipts would not be counted as income in any of the other senses in which the term income is used.

We must also exclude from the income total we seek anything for which neither a good nor a service is supplied in exchange and for which there is therefore no corresponding production. One example is the billions of dollars received by persons who sell stocks and bonds they own—these are simply financial transactions, and what the sellers receive is, again, neither "income" in the sense of something generated in the course of producing output during the time period nor is it even income in any of the other senses usually given to this term. A less obvious example of the same sort of exclusion is the income received by persons from goverment in the form of transfer payments, i.e., income for which the recipient provides no good or service in exchange. Just as government transfer payments had to be excluded on the expenditures side because they are not expenditures for a good or service, so they must be excluded on the income side because they are not income to which there corresponds production. The same is true for income received by persons from other persons for which no productive service is rendered. These are nothing more than interpersonal transfer payments and as such must be excluded from the income total that is being sought. That income total is to be a measure of the value of the economy's output of final goods and services; it

can only be such a measure if it excludes every dollar of income which is not matched by a dollar's worth of production of final goods and services. The flows here marked for exclusion are clearly flows unmatched by any production of goods and services.

What Counts as Income? What forms of income then do qualify in this way? It might seem at first glance that we can include under this heading all of the receipts of business firms because each dollar of their receipts or their "income" is matched by a dollar's worth of goods or services sold. However, it should be apparent that adding up all these amounts will involve the same sort of double, triple, or higher multiple counting that we saw must be avoided on the expenditures side. We may no more count all the receipts of the wheat farmers, the millers, and the bakers as part of the income total being sought than we could count all the spending by millers, bakers, and persons as part of the expenditure total being sought. Recall that only the spending for bread is spending for final product, and the baker's spending for flour used up in making bread and the miller's spending for wheat used up in making flour which in turn is used up in making bread are spending for intermediate product. If a loaf of bread is purchased by a person for 60¢, that is the only spending to be counted in connection with this output.

In working on the income side, we seek a total for income that will be equal to the value of the economy's output of final goods and services. Accordingly, in the present illustration, income must be defined so that there is a total of 60¢ of income, no more and no less, generated by the production of a loaf of bread. Clearly, if the method of counting income does indeed give a total of 60¢ of income for the loaf of bread, the approach on the income side yields the same result as the approach on the spending side. We will get this same result, 60¢ of income to accompany the loaf of bread, if from the *total receipts* of the baker, the miller and the farmer is deducted whatever each of these paid to purchase materials from other firms that each in turn used up at his stage of production. This is the equivalent of the earlier discussed process of cancelling out intermediate product to arrive at final product, but here it is viewed as a cancelling-out of part of a firm's total receipts to arrive at the part that may be called income in the sense here given to that term.

For a simple numerical illustration, suppose again that a consumer spends 60¢ at the bakery shop for a loaf of bread. As we have seen, this is spending for final product and that 60¢ enters into the total that is GNP for that period. The baker obviously shows 60¢ of receipts as a result of this transaction, but that 60¢ is not all income as the term is here defined. From it must be deducted the amount of materials purchased from other firms. Assume for simplicity that the only material needed to make bread is flour and that the purchase price of the amount of flour that goes into a loaf of bread is 32¢. Deducting 32¢ from 60¢ means that 28¢ of the 60¢ secured by the baker is to be counted as income. Assume that the miller who received the 32¢ from the baker paid 20¢ to the farmer for the amount of wheat involved—deducting the 20¢ from the 32¢ means that 12¢ of the 32¢ received by the miller is to be counted as income. Lastly, assuming that the wheat farmer makes no purchases from any other firms, there is no deduction from the 20¢ he receives so that it is to be counted as income in its entirety. Summing up, we find that the loaf of bread is matched by 20¢ of income generated at the farmer stage, 12¢ of income generated at the miller stage, and 28¢ of income generated at the baker stage, or a total of 60¢, the same amount at which this unit of output would be valued when approached from the expenditure side.

Although our purpose here is to focus on the fact that the 20¢, 12¢, and 28¢ amounts are amounts of income generated at the successive stages of production, it should be noted that these amounts are also commonly described as

amounts of *value added*. Thus, because the miller takes 20¢ of wheat and turns it into something he sells for 32¢, he is said to add 12¢ of value to the wheat. In the same way, the baker adds 28¢ of value to the flour purchased from the miller or the difference between 60¢ and 32¢. Because the farmer did not purchase anything from any other firm, the entire 20¢ for which he sells the wheat is value added.

The term value added may seem more appropriate than the term income to describe the amounts in question. Most people think of the amount of income originating in a bakery or on a farm or in any other firm as the amount of the firm's *profit*. Here income is used in a way that includes profit and much more. To turn again to our illustration, in the case of the baker we know that 32¢ of materials (assuming flour to be the only material) are used up in producing a loaf of bread that can be sold for 60¢. However, the remaining 28¢ which is here called income is not the profit of the baker on this unit of production. Suppose that 20¢ is the amount of wage and salary costs incurred in producing this unit of output, that 2¢ is due the government for sales tax included in the sale price, that 1¢ is to be allowed to cover depreciation of the baker's ovens and other durable assets, and that another 1¢ is charged to meet the interest cost on the firm's debts. If anything is left, that is the amount of the baker's profit on this unit of output. The present illustration shows this residual to be 4¢, but it could be a larger or smaller amount, even a negative amount, depending on the total of the other costs. The baker is, for obvious reasons, primarily concerned with what this profit residual is, but we are primarily concerned with finding the total amount of income that originates in each firm as a step toward deriving an estimate for the value of the economy's output as measured from the income side.

It should be clear that what was described here for the baker could in the same way be described for the miller and the farmer. In each case, there is no problem in seeing why the wage and salary costs incurred by each firm are called income, once one looks at these from the viewpoint of the worker to whom they are receipts. Similarly, the firms' interest costs, indirect tax costs, and the residual called profit are all plainly income from the viewpoint of the recipient. Depreciation costs are also a part of income by the present definition, but a part retained by the firm and in effect set aside for possible future use. In this way, corresponding to the 60¢ value of the loaf of bread will be found 60¢ of income made up of wage and salary income, interest income, profit income, government income from sales and related types of taxes known as indirect taxes, and an amount of income retained by firms in the form of additions to depreciation reserves.[4]

Once having established what is to be included as income, we need not concern ourselves with the income generated by the production of any one loaf of bread or by all loaves of bread combined or even with the income generated by the whole baking industry. We may turn directly to industry as a whole. The immediate purpose is to estimate from the income side the value of the whole economy's output of final goods and services and one may proceed straightforwardly toward this objective by preparing economy-wide estimates of the totals for wages and salaries, interest, business profits, and other income items that are to be included. Apart from a few minor items not here entered into, the total of these amounts is the total income of the economy in the present sense of the term. Once allowance is made for the items not here covered, the total so obtained will be the same as the total that was reached on the expenditures side, assuming equally good data and equally good estimating on the two sides.

[4] Government will, of course, take part of wage and salary income, interest income, and profit income through direct taxes or income taxes, but that is another matter—here we look at these income receipts before income taxes on them are paid.

National Income, Personal Income, and Disposable Personal Income Recall that the approach on the expenditures side yielded two totals, GNP and NNP, which differ by the amount of capital consumption. We could identify two corresponding totals on the income side and assign corresponding names to them. Thus, some people choose to label as Gross National Income the total on the income side that is equal to GNP on the product side. Similarly by subtracting capital consumption from Gross National Income on the income side, the remainder may be called Net National Income since it equals Net National Product on the product side. However, usual practice is to use only the GNP and NNP terms when referring to these totals. As shown in Table 2-1 which, in general, follows the National Income and Product Account as presented by the Department of Commerce, the two totals in question are not designated income totals but Charges against GNP and Charges against NNP.

GNP and NNP are two of three measures of the economy's output presented in the U.S. national income accounts. National Income, NI, found on the income side of Table 2-1 is the third measure of output. NI values output at "factor cost" or in terms of the factor incomes earned (though not necessarily received) by the factors of production. In the classification shown on the income side of Table 2-1, NI is the sum of compensation of employees, proprietors' income, rental income of persons, corporate profits, and net interest. Each of these income flows is a return for the amount of a service rendered by a factor of production, and the total is the amount earned by all the factors of production for all factor services rendered by them. That total is, therefore, a measure of the value of the economy's output based on factor cost. On the expenditures side, GNP and NNP are derived by adding up expenditures on goods and services, with each unit of each good and service valued at the price paid for it. This approach values goods at "market price," and the market price of any good from a loaf of bread to a jet aircraft will typically be greater than the "factor cost" of producing that good. The market price will ordinarily include an amount to cover depreciation of the capital goods used in its production and an amount to cover any indirect taxes, e.g., sales taxes, that must be paid by its seller to the government.

As we have seen, NNP for any period is a measure of a smaller amount of actual output than is GNP, because NNP excludes the portion of output equal to the capital goods used up in producing that period's output. NNP and NI, however, are both measures of the same amount of goods, although they value these goods on different bases. NNP will always be greater than NI, because the market price of the amount of goods in question will always exceed the factor cost by approximately the amount of indirect taxes. Thus, as shown by Table 2-1, for 1976 the NNP of $1,527.4 billion exceeded the NI of $1,364.1 billion by $163.3 billion, which is almost entirely accounted for by indirect taxes of $150.5 billion and business transfer payments (gifts, prizes, and the like for which business receives no good or service in exchange) of $8.1. Two minor items make up the balance of the difference between NI and NNP: subsidies less current surplus of government enterprises and the statistical discrepancy.[5] The latter is the difference between the estimated sum of the items on the product side and the items on the income side. If the sum on the income side exceeds that on the product side, a negative figure of that amount is entered on the income side as the statistical discrepancy to make the sum on that side equal the sum on the product side; in the opposite case, the appropriate positive amount is entered.

Although they are not measures of the economy's output, there are two other important income measures, Personal Income, PI, and Disposable Personal Income, DPI, the second of which is especially important for our later

[5]The first of these items is somewhat technical in nature; a brief explanation may be found in the appendix, pp. A-28.

work on personal consumption spending. Personal Income is the current income of persons from all sources. It is not a measure of output or production because it is a total that includes some amount to which there corresponds no production. Thus, it includes both receipts for the productive services provided by persons and receipts, such as transfer payments, for which no productive services were provided by the recipients. DPI is derived from PI simply by deducting from PI the amount taken by government in personal taxes. The remainder is available to persons to dispose of as they see fit. Because DPI is derived from PI, which is not a measure of output, DPI in turn is not a measure of output.

The way in which PI and DPI are related to the three measures of output—GNP, NNP, and NI—and also the way in which these measures of output are related to each other is best seen by starting with the broadest aggregate, GNP, and identifying the flows that are subtracted and added in moving from one aggregate to the next. This is done in Table 2-2. As in Table 2-1, the figures are for 1976.

The steps downward from GNP to NNP and

TABLE 2-2
Relation of gross national product, net national product, national income, personal income, and disposable income, 1976
(billions of dollars)

Gross National Product	1,706.5			1,706.5	
less Capital consumption allowances with capital consumption adjustment	179.0			is sum of	
Net National Product	1,527.4				
less Indirect business taxes	150.5				
Business transfer payments	8.1				
Statistical discrepancy	5.5		S	278.0	
plus Subsidies less current surplus of government enterprises	0.8	(−)			
National Income	1,364.1				
less Corporate profits taxes	64.7				
Contributions for social insurance	123.8				
Undistributed corporate profits with inventory valuation and capital consumption adjustments	27.6				
Wage accruals less disbursements	0.0	(−)			
plus Government transfer payments to persons	184.7		T	333.5	0.1 off due to rounding
Net interest paid by government to persons	16.9	(−)			
Interest paid by consumers	25.0				
Business transfer payments	8.1				
Personal Income	1,382.7				
less Personal taxes	196.9				
Disposable Personal Income	1,185.8				
less Personal saving	65.9				
Interest paid by consumers	25.0				
Personal transfer payments to foreigners	0.9		R_{pf}	0.9	
Personal Consumption Expenditures	1,094.0		C	1,094.0	

SOURCE: *Survey of Current Business*, U.S. Department of Commerce, July 1977.

from NNP to NI in Table 2-2 will be seen to be the same as the steps upward from the total at the bottom of the income side of Table 2-1 to the NI figure on that side of Table 2-1. However, whereas Table 2-1 then shows above NI the various factor income flows which add up to NI, Table 2-2 proceeds to show how another income aggregate, PI, is related to NI. First, from NI are subtracted those portions of NI that are not received by persons. Thus, as shown in Table 2-1, the major component of NI is compensation of employees and the figure for this component is an estimate of the amount earned for the labor services provided by all employees. However, if we want to know the amount actually received by all employees for their labor services, we must deduct from compensation of employees the contributions for social insurance which is an amount diverted to the government.[6] Similarly, the total of corporate profits is factor income and thus part of NI, but part of this total is paid to government in corporate profits taxes and part is withheld by corporations as undistributed profits. Only the remainder that is dividends is passed on to become income of persons or, in other words, only this part of corporate profits is included in PI. The treatment of the other components of NI, namely proprietors' income, rental income of persons, and net interest, shows the full amount of each passing on to become income of persons.

Next, to arrive at the total for PI, it is necessary to add to the portion of NI that becomes PI other income of persons that is not income received for productive services rendered. As shown by Table 2-2, the additions are government transfer payments, interest income from government and from consumers, and business transfer payments, with the first of these items being by far the largest. It will be recalled that only government purchases of goods and services are included on the expenditure side of Table 2-1 as there is a dollar of production corresponding only to each of these dollars of government spending. Therefore, this portion of government spending is included in GNP and thus is part of the total at the top of Table 2-2. The balance of government spending which is primarily made up of government transfer payments is not part of GNP but is part of PI and is picked up in reaching the PI total by inserting it in Table 2-2 as shown. The other major item that is not part of GNP but is part of PI is certain interest income. Briefly, the net interest component of NI in Table 2-1 is interest paid by business. This kind of interest is treated as part of NI because business uses the borrowed funds for productive purposes. Borrowing by government and by consumers is not viewed as borrowing for productive purposes and the interest paid on such borrowing is, for this reason, not included in NI. However, like government transfer payments, this interest is counted as part of PI and is inserted as shown to arrive at that total. Finally, there is the addition of the amount of business transfer payments. This amount is part of GNP but is subtracted in moving from GNP to NI because such payments are not factor income to their recipients who provide no productive service in exchange. However, they are income to persons and are added back in to get from NI to PI.

After working down to PI from GNP, the step from PI to the final income aggregate, DPI, is made by subtracting personal tax payments such as personal income taxes from PI. DPI is then the amount available to persons to dispose of as they choose, and they choose to save some amount of this total each year. Deducting from DPI the amount of personal saving and amounts for personal transfer payments to foreigners and interest payments by consumers leaves the amount of personal consumption

[6] One might argue that a deduction should be made for the amount of income taxes withheld from paychecks because that amount of compensation of employees also is not received by persons. However, as this is only part of the income taxes paid by persons, the practice is to include as Personal Income all such income before the personal income taxes thereon and then deduct the total of such income taxes in one step going from PI to DPI.

expenditures, C, which makes up the great bulk of DPI every year. The figure found here for C, it will be seen, is the same figure shown for that flow on the product side of Table 2-1.

ACCOUNTING IDENTITIES

The structure of Table 2-2 has permitted us to start out with a figure for the GNP or the gross income flow and to work down from this gross figure to find how much of it was spent on consumer goods and services. Thus out of the gross of \$1,706.5 billion for 1976, \$1,094.0 billion was spent on consumption. The structure of Table 2-2 also permits us to derive the amount of the gross income flow that was taken by government in taxes and the amount that went into private saving. Given the way that saving and taxes are defined in the accounts, it will be found that the difference of \$612.5 billion between the gross income flow and the amount of it spent on consumption is, apart from a minor amount for personal transfer payments to foreigners, entirely accounted for by the sum of taxes, \$333.5 billion, and private saving, \$278.0 billion.

Bringing together the breakdown on the product side that was earlier introduced and the breakdown on the income side here presented gives the following GNP identity:[7]

$$
\begin{aligned}
C \quad + \quad S \quad + \quad T \quad + R_{pf} &= \text{GNP} \\
\$1{,}094.0 + 278.0 + 333.5 + 0.9 &= 1{,}706.5 \\
= C \quad + \quad I \quad + \quad G \quad + &\ (X - M) \\
= 1{,}094.0 + 243.3 + 361.4 + &\ (162.9 - 155.1)
\end{aligned}
$$

in which S is the sum of personal saving and business saving and T is the sum of social insurance contributions or taxes, indirect taxes, corporate profits taxes, and personal taxes after an adjustment of that sum that will be noted below. The derivation of the figures in this identity may be seen by following the lines in Table 2-2 which classify various items listed on the left under C, S, T, and R_{pf} on the right.

The figure for C appears explicitly as Table 2-2 is structured to show this figure as one end product. The minor item, R_{pf}, is also explicitly identified. The figure for S is derived by adding together the amounts of the gross income flow saved by businesses and by persons, the private domestic sectors of the economy. This total is found as the sum of the amounts given for four items. First, the amount of the gross income flow that shows up as capital consumption allowances, \$179.0 billion, is a kind of business saving, a portion of the gross flow received by businesses that is set aside for replacement of capital goods. Next, the amount of the gross flow that is retained by businesses as undistributed profits, \$27.6 billion, is quite clearly saving, a portion of corporate profits that are not paid out to anyone else.[8] The sum of these two amounts \$206.6 billion, is total business saving. The total for private saving, the sum of business and personal saving, is then obtained by adding to the figure for total business saving the amount of personal saving, \$65.9 billion, and the minor amount for the statistical discrepancy, \$5.5 billion (which must be

[7]To simplify I now appears without a subscript. However, if reference is to GNP, one knows that I must be I_g, and if reference is to NNP, one knows that I must be I_n. Similarly, on the income side, if reference is to the gross concept, S must include the capital consumption allowances component of private saving, and if reference is to the net concept, S must exclude this component of saving.

[8]The original amount of undistributed profits, \$56.4 billion, is adjusted downward by \$28.8 billion, the sum of the inventory valuation adjustment and the capital consumption adjustment. These adjustments are explained in the appendix, pp. A-56–A-59.

Net Exports
Goods & Services

put in somewhere to provide a balance and is put in here on the assumption that the discrepancy arises from errors in the estimates of the indicated saving flows). Thus, S of \$278.0 billion emerges as the sum of \$206.5 billion of business saving, \$65.9 of personal saving, and \$5.5 for the statistical discrepancy.

How much of the gross income flow is taken by government in taxes? In Table 2-2, four of the lines running to T show the four kinds of tax receipts of government into which all tax receipts are classified. As shown in the table, these are indirect business taxes, \$150.5 billion, corporate profits taxes, \$64.7 billion, contributions for social insurance, \$123.8 billion (a tax in effect though not in name), and personal taxes, \$196.9 billion. The total is \$535.9 billion. It will be seen that if we add this total for T to the amounts already found for C, S, and R_{pt}, the sum will be \$1,908.8 billion, which is \$202.3 billion greater than GNP or, what is the same thing, that much greater than the total for $C + I + G + (X - M)$. The discrepancy arises from the fact that we quite correctly include only government purchases of goods and services in G on the product side and so far have included gross tax receipts on the income side. To be consistent, as all government expenditures other than those for goods and services are omitted on the product side, an equal amount of tax receipts should be omitted on the income side. To show tax receipts on such a net basis, an amount equal to government expenditures other than those for goods and services is deducted from gross tax receipts. These expenditures are identified in Table 2-2 by the other three lines running to T: subsidies less current surplus of government enterprises, \$0.8 billion, government transfer payments, \$184.7 billion, and government interest payments, \$16.9 billion. Subtracting the sum of these three items, \$202.4 billion, from gross taxes of \$535.9 billion leaves net taxes of \$333.5 billion, which is the net tax concept to which T in Table 2-2 refers. With taxes on a net basis, the sum of $C + S + T + R_{pt}$ is found to be identical with GNP.

As C appears on both sides of the GNP identity above, another basic identity is obtained by dropping C from both sides.

$$\begin{aligned} S \quad + \quad T \quad + R_{pt} &= \quad I \quad + \quad G \\ \$278.0 + 333.5 + 0.9 &= 243.3 + 361.4 \\ &\quad + \quad (X \quad - \quad M) \\ &\quad + (162.9 - 155.1) \end{aligned}$$

The sum on the left is the amount of the gross income flow not devoted to consumption expenditures and the sum on the right is the amount of the gross output flow that does not go into the hands of consumers. Roughly, this identity says that during any time period there is a dollar of saving or taxes in the system for every dollar of investment, government purchases, and net exports. The full significance of this identity and the GNP identity above will become clear in Part 2 where these accounting relationships provide a framework within which to build the theory of income determination.

The GNP identity above is one which includes all four sectors—consumers, businesses, governmental units, and the rest of the world—found in the actual economy. However, in developing the theory of income determination in Part 2, we will not start off by including all four sectors and we will therefore not start with the GNP identity for a four-sector economy. To approach the complications of the real economy by steps, we begin with a hypothetical economy limited to consumers and businesses only. In Chapters 4 and 5, there are neither government spending nor taxing and there are neither exports nor imports (nor foreign transfer payments). It is apparent that in such a two-sector economy the GNP identity above is reduced to

$$C + S = \text{GNP} = C + I$$

The omission of government removes G and T, and the omission of the rest of the world removes $(X - M)$ and R_{pt}. On the product side, the amount of the economy's output obtained by consumers is measured by personal consumption expenditures, and the amount that remains is investment for that time period. On

the income side, the income flow equal to the value of output is divided between personal consumption expenditures and saving.

Another identity for the two-sector economy is derived from the first by dropping C from both sides.

$$S = I$$

For the simple case of a two-sector economy, this identity says that there is a dollar of investment for every dollar of saving. Because there is a dollar of income for every dollar of output and because the amount of income not spent for consumer goods is equal to saving in this two-sector economy, it follows that the amount of saving is equal to the amount of investment. Or saving, the amount of income that is not used to purchase consumer goods and services, is by definition equal to investment, the amount of output that does not go into the hands of consumers.

A special property of a two-sector economy is that NNP, NI, PI, and DPI all become equal if we make a few assumptions. With government excluded, most of the items that are found between NNP and DPI in Table 2-2 simply drop out. If we assume that corporations pay out all earnings in dividends, i.e., undistributed profits are zero, and assume a zero value for the statistical discrepancy, business transfer payments, inventory valuation adjustment and capital consumption adjustment, and consumer interest payments, the result will be NNP = NI = PI = DPI. This special case in which there is a dollar of DPI for every dollar of NNP is the one that is employed in the analysis of Chapters 4 and 5.

The three-sector economy which is examined in Chapter 6 adds the government sector to the consumer and business sectors but excludes the rest-of-the-world sector. There is then a three-way split on the product side into $C + I + G$ and on the income side into $C + S + T$ or

$$C + S + T = \text{GNP} = C + I + G$$

As before, by dropping C from both sides, another identity, a saving and investment identity, is derived.

$$S + T = I + G$$

which may also be written as

$$S + (T - G) = I$$

S has been identified as private saving, the sum of consumer and business saving. Total saving is the sum of private saving and public saving, and public saving is $(T - G)$, positive for a budget surplus and negative for a budget deficit. With government included, there is still an identity between saving and investment, as now the sum of private and public saving is by definition equal to investment. However, unlike the identity for the two-sector economy, private saving or S is no longer by definition equal to I. S will equal I only if $(T - G)$ equals zero, i.e., only if there is a balanced budget. With a balanced budget, there is neither public saving, a surplus, nor public dissaving, a deficit.

Finally, bringing in the rest-of-the-world sector returns us to the four-sector economy with which we started and for which the GNP identity is

$$C + S + T + R_{pf} = \text{GNP} = C + I + G + (X - M)$$

Chapter 7 takes up the four-sector economy and it is this identity with which we will work in that chapter.

Chapters 4–7 are concerned with the development of what is commonly called the simple theory of income determination, and the preceding series of identities that have emerged from the national income accounting covered in this chapter will be used in those chapters in the order indicated. However, before turning to those chapters which make up Part Two, it will be helpful to examine a few of the basic concepts whose understanding is essential to the theoretical material in Part Two and the following parts of the book. This is the task of the following chapter.

chapter 3

Basic Concepts

The U.S. economy's output of goods and services during 1976 was $1,706.5 billion as measured by GNP. What determines the size of this amount for any year? This question defines one of the major tasks of macroeconomic theory and one that will be the major concern through Parts Two, Three, and Four of this book. It also leads logically to further questions: What determines the fluctuations in output from one year to the next (which is essentially the question of the business cycle)? What determines the rate at which output increases over a period of years (which is essentially the question of economic growth)? We will look into these additional questions in Part Five.

The practical importance of the first question we raised cannot be exaggerated, for its answer provides in part the answer to the fundamental question of what determines the material well-being of the economy's population. Although qualified by the need to protect the environment and to avoid the depletion of nonreproducible resources, a short-run goal of any economy is the production in each year of the maximum amount of goods and services possible with its currently available labor force, stock of plant and equipment, and technological know-how. An economy that managed to maintain such a maximum flow of output would thereby be providing its population with the highest per capita real income attainable each year. In an economy that failed to achieve this goal, the per capita real income of its citizens would be below what it could be; moreover, some people who wanted to work, and who could thereby increase aggregate output and per capita real income, would find themselves without jobs. When an economy fails in any period to produce actual output equal to its potential for that period, we cannot begin to explain its failure—or formulate a policy to remedy it—unless we can first explain what determines its actual output. In other words, whether actual output is at potential or below potential, to explain what we actually observe, we must have a theory of what determines aggregate output.

If the economy's actual output in every year were equal or nearly equal to its potential for that year, however, the theory of the determination of aggregate output would be a far less important subject than it is. Attention would then focus primarily on growth theory, which attempts to explain the rate at which aggregate output grows over time. If that growth rate were, in turn, one that provided a very rapidly rising standard of living as well as uninterrupted full

utilization of the labor force and other resources, neither the short-run theory of the determination of aggregate output nor the long-run theory of economic growth would be as important as they are. Unfortunately, instead of these ideal results, the actual record for the United States, as well as for other economies, shows sizable fluctuations in the level of output with attendant underutilization of labor and other resources and also, in some periods, a growth rate that has meant a slowly rising or even falling per capita real income. In view of this actual record, the short-run theory of income determination and the long-run theory of economic growth are of much more than academic interest. They involve the problem of securing the highest possible standard of living for the nation's people, and there is probably no more important problem within the province of economics.

FROM MACROECONOMIC ACCOUNTING TO MACROECONOMIC THEORY

Although it may seem otherwise at first glance, national income accounting does not provide the answer to what determines the economy's actual level of output in any period. The data provide us with estimates of the economy's output for any time period, but they do not tell us what determines that output or what determines the changes that occur from one time period to another. For example, we may compare GNP in constant dollars for two years and label the difference the change in the economy's output from the first to the second year.[1] Thus, for the U.S. economy in 1975 and 1976 we have the following figures in billions of 1972 dollars for the product side of the GNP identity:

$$
\begin{array}{lrcl}
 & \text{GNP} & = & C + I + G + (X - M) \\
1975 & 1{,}202.1 & = & 775.1 + 141.6 + 263.0 + 22.5 \\
1976 & 1{,}274.7 & = & 821.3 + 173.0 + 264.4 + 16.0
\end{array}
$$

The figure for GNP in 1976 is simply the official estimate of the final output flow for the year, and the figures for C, I_d, G, and $(X - M)$ are simply estimates of the composition of this total flow. GNP of $1,274.7 billion was necessarily identical with the sum of its component parts, since it was derived by summing these parts.[2] Although this is valuable for other purposes, it is nothing more than an identity and as such is valueless as an explanation of *why* output was actually this amount in 1976.

What then did determine the economy's output in 1976 or in any other period we might choose to consider? As a first step toward any kind of an answer to what is a very complex question, we plainly have to discover the major variables that influence how much output the economy produces and to detect the relationships among these variables that give rise to the actual results we find in any period. In other words, as a first step we need a theory of income determination, which, in full-blown form, is a detailed analytical framework or model that expresses in the various relationships that make up that model how each variable is believed to be related to the other variables that have been identified as relevant to the problem. Such relationships are functional relationships in the

[1] Readers not familiar with the distinction between current dollar and constant dollar GNP will find an explanation in the appendix, pp. A-59–A-62.

[2] GNP may be broken down in other ways without altering this conclusion. For example, by major type of product we have durable goods of $235.2 billion, nondurable goods of $344.9 billion, services of $584.7 billion, and structures of $109.9 billion, all adding up to GNP of 1,274.7 billion in 1976. In no sense, however, does this alternative breakdown—or any other—tell us what actually determined GNP in 1972.

sense that one variable in the relationship is believed to be a function of one or more other variables in a way specified by the particular theory that has been advanced. The relationship that makes aggregate consumption expenditures a function of disposable personal income is such a relationship; it expresses a theory, however simple, of the determination of aggregate consumption. It is only by devising theories, by hypothesizing functional relationships, that we can make any progress toward explaining the facts revealed, for example, by the data in the preceding identities. And devising theories, however indispensable in the process of explanation, is only a step in the process, for the theories devised may or may not be supportable. In order to decide which theories are to be at least tentatively retained and which are to be rejected, we must test them against the "facts." Furthermore, a theory that is so supported is only provisionally accepted, for such support can never "prove" a theory. Sometimes we find a number of conflicting theories that receive equal support from the available data, and the question of which is the "true" theory remains at least temporarily unresolved.

We can bypass these and other complications that are faced in actually constructing and testing a detailed model and still say something meaningful about what determines the economy's output by dropping down to the simplest possible model of Keynesian theory. In that theory the basic force determining aggregate output is aggregate demand, and the simplified model that emerges includes little more than a few functional relationships designed to explain in turn the determination of the level of each of the major components into which aggregate demand may be divided. There is a theory for aggregate consumption expenditure (the one noted above) and an equally simple theory for the other components. Nobody who seriously sets out to explain output determination limits himself to so crude a model, but even here we have a model that, despite its simplicity, takes us at least some distance toward answering the question we raised.

Unlike the simplest theoretical framework, no accounting framework, however elaborate, can in itself provide the answer we seek. An accounting framework is based on identities, and identities in themselves do not explain. The accounting framework and the definitions of the items that comprise it are nonetheless essential as a setting within which may be developed the theoretical framework that will give us the answer. The more detailed the theoretical framework, the more detailed must be the accounting framework that supports it. If our purpose were only to develop the theory of the determination of income and output for a hypothetical two-sector economy, all we would need would be the simple accounting framework for such an economy from which emerge the fundamental identities, $C + S = \text{GNP} = C + I$ and $S = I$. If our purpose were to develop the theory for a hypothetical three-sector economy, we would need the accounting framework for such an economy from which emerge the fundamental identities, $C + S + T = \text{GNP} = C + I + G$ and $S + (T - G) = I$. And if the purpose were to do the same for a hypothetical four-sector economy, we would need a corresponding accounting framework from which emerge the fundamental identities, $C + S + T = \text{GNP} = C + I + G + (X - M)$ and $S + T = I + G + (X - M)$. (In a hypothetical four-sector economy, one may simplify by assuming $R_{pf} = 0$). Beyond this, if our purpose were to use the developed theory as the basis for policy, and such is the ultimate purpose of theory, we would need the detail of a real-world, four-sector accounting framework such as the one provided by the U.S. Department of Commerce that is presented in the appendix. And, finally, if our purpose were to test the developed theory against the facts, and such is obviously desirable for any theory, we would again need a real-world accounting framework and the "facts" for the real world as provided in that framework. The relationships between the accounting

framework and the theoretical framework present a question that will be answered in the later parts of this book. We will see how macroeconomic theory, and especially the application of this theory to questions of policy, is intimately tied to macroeconomic accounting.

Before plunging into the development of macroeconomic theory, it will be helpful to examine a few of the basic concepts that run the full breadth of economic theory, such as stocks and flows, equilibrium and disequilibrium, and statics and dynamics. The presentation here is designed to provide only an elementary understanding of what are actually some of the most troublesome concepts in economic methods.[3] Furthermore, the meaning assigned here to each of these concepts is only one of the several meanings that economists have employed. Finally, the coverage in what follows is limited to the three pairs of concepts noted above; the many other concepts of narrower application may be more advantageously treated, where pertinent, in the parts ahead.

STOCKS AND FLOWS[4]

The twin concepts of stocks and flows are not especially difficult to understand, but they can cause great difficulty if misunderstood or misused. To begin with, stocks and flows are both variables; they are quantities that may grow smaller or larger over time. The distinction between them is that a *stock* is a quantity measurable at a specified *point* in time and a *flow* is a quantity that can be measured only in terms of a specified *period* of time. For example, a gauge may indicate that the stock of water in a reservoir is 50 million gallons; the stock variable is 50 million gallons at this particular point in time. It would be meaningless to describe this as 50 million gallons a year, a month, a week, or a day. Another gauge may indicate that the flow of water into the reservoir amounted to 365 million gallons over the year then ended. Assuming that the flow was at a fixed rate over the year, this reading would also indicate that water had flowed in at a *rate* of 7 million gallons per week or 1 million gallons per day.

As another example, consider the total number of persons employed in the United States—this is a stock variable. In contrast, the number of persons who secure new jobs or leave employment are flow variables. The number employed is, say, 90 million at a point in time (on a particular day); it is nonsense to speak of the number employed as 90 million per year.[5] The number of persons who find employment may be, say, 100,000 for a given time period, the month of June. This is not 100,000 at a specific point in time, however.

Money is a stock, but the spending of money is a flow. To say simply that the stock of money is $300 billion has no meaning until we specify the point in time—March 31, 1976—at which this was the stock.[6] Similarly, the statement that total spending for final output amounted to

[3] For an advanced discussion of these particular concepts, see J. R. Hicks, *Capital and Growth*, Oxford Univ. Press, 1965, Chs. 1–3 and 8.

[4] See also G. Ackley, *Macroeconomic Theory*, Macmillan, 1961, pp. 5–8, and W. S. Vickrey, *Metastatics and Macroeconomics*, Harcourt Brace Jovanovich, 1964, pp. 116–17.

[5] We can and do say that employment or the number of persons employed *averaged* 90 million during the year, a figure derived by estimating employment at a number of specific points in time (e.g., the middle of each month during the year) and then computing the average value of these estimates. The average figure, no less than each of the twelve mid-month figures from which it was derived, is still a stock variable. This is not to be confused with the "unemployment rate," however, which is a *ratio* of two *stock* variables, the number unemployed divided by the number in the labor force.

[6] As in the case of employment, we can say that the stock of money averaged so many billions of dollars for the year, but again this average figure is a stock variable.

$1,706 billion is meaningless until we specify the time period, the year 1976, during which this amount was spent. Here we can see the serious errors that can result from a failure to distinguish stocks from flows. Some people fail to make a distinction between the *amount* of money and the *spending* of money. They simply equate the two, perhaps because whatever money they get their hands on they promptly spend. From this error follows the more serious error of imagining an increase in the stock of money to be a certain means of producing an equal increase in the flow of spending. Far from being equal, however, the two can and at times do change in opposite directions to produce a combination of more money and less spending or less money and more spending. As soon as it is recognized that the variable money is a stock and the variable spending is a flow, there can be no equating of the two.

There are other illustrations of the stock/flow distinction in the national income and product account on p. 19. Every entry in that account is without exception a dollar figure measuring a flow. Some of these figures, such as "change in business inventories," may at first glance appear to measure stocks. Notice, however, that the entry is not "inventories," which is clearly a stock, but rather "*change* in inventories," which is just as clearly a flow, for a change in any variable can only be measured over a period of time.

Some macroeconomic variables that have flow magnitudes also have direct counterpart stock variables. However, others, such as imports and exports, wages and salaries, tax payments, social security benefits, and dividends, are only flows; none has a direct stock counterpart as it is impossible to conceive of a "stock of imports" or a "stock of wages and salaries." Although such flows have no direct stock counterpart, they do indirectly affect the sizes of other stocks. Imports may affect the size of business inventories or the stock of capital goods; wage and salary receipts devoted to the purchase of newly produced houses may affect the stock of housing. In the case of some flows that have a direct counterpart in a stock, statistics on both the stock and the flow variable are unfortunately reported under headings that are practically the same. A person's saving is a flow ($25 for April), and his saving*s* are a stock ($500 accumulated as of April 30); a firm's gross investment is a flow ($50,000 for April), and the *total invested*, or the dollar value of real capital accumulated, is a stock ($1 million as of April 30); the *change* in the nation's money supply is a flow ($1 billion increase during April), whereas the money supply itself is a stock ($300 billion as of April 30).

For those flow variables that have a direct stock counterpart, any change in the magnitude of the stock variable between two specified points in time depends on the magnitudes of its counterpart flow variables during the period.[7] Thus, the number of persons employed increases, decreases, or remains unchanged between two points in time, depending on the number of persons who secure employment and the number of persons who leave employment during the intervening period. The nation's stock of capital changes between two points in time depending on the inflow (the amount of gross investment or capital goods produced) and the outflow (the amount of capital goods consumed) during the intervening period.

Although it is necessarily true that a stock can change only as a result of flows, the magnitudes of the flows themselves may be determined in part by changes in the stock. The best example is the relationship between the stock of capital and the flow of investment. The stock of capital can increase only as a result of an excess of the flow of investment or of new capital goods pro-

[7]Since stocks and flows in macroeconomics are usually expressed in dollars, a change in a stock may occur with no change in the real counterpart flows but simply as a result of a change in the basis of valuation for a given physical stock. Thus, the existing U.S. gold stock increased in value in 1972 and again in 1973 with the devaluations of the dollar in those years. Also, since August 15, 1971, changes in the monetary gold stock via outward flows of gold ceased with the suspension of convertibility as of that date.

duced over the flow of capital goods consumed. However, the flow of investment itself depends, among other things, on the size of the capital stock. In many theories of the business cycle, a critical factor in the explanation of business downturns is a decrease in the flow of investment brought on by an "excessive" stock of capital resulting from an earlier, prolonged upsurge in the flow of investment. This earlier upsurge in the flow of investment was usually brought on by a decrease in the stock of capital during the preceding depression, a period during which the flow of investment fell below that of the preceding period of prosperity. As is apparent, this process may continue ad infinitum and carry with it the endless sequence of ups and downs known as business cycles.

By definition, stocks can exert an influence on flows only if the time period is long enough to produce the required change in stocks. Where stocks are very large relative to flows, the changes in stocks resulting from flows are typically so small in the short-run period that stocks may be assumed to be constant in that period. Thus, although flows may be influenced by changes in stocks, it follows that they will not be so influenced by changes in stocks in the short run. For example, if the net effect of the flows of gross investment and capital consumption is an increase in the stock of capital between January 1 and December 31 amounting to a fraction of 1 percent of the January 1 stock of capital, then the capital stock may be assumed to be approximately constant. Since it is approximately constant, it can have no significant effect on the flow of net investment in the following period.[8]

With respect to this relationship between the flow of investment and the stock of capital, we may define the short-run period as one in which changes in the stock of capital are too small to have an influence on the flow of investment and the long-run period as one in which such changes are large enough to have an influence on the flow of investment. In this sense, elementary macroeconomic theory is primarily short-run; it is essentially a study of relationships among flows in which the size of each flow in any time period is determined solely by the sizes of other flows. In the simplest formulation of Keynesian theory, the flow of consumer spending is determined by the flow of income, and the flow of income equals the flow of consumer spending plus the flow of investment spending. Although we are primarily concerned with elementary theory in this book, we will devote some attention to more advanced theory in which changes in such critical stocks as the stock of money and the stock of capital affect the all-important flows of income and product.

EQUILIBRIUM AND DISEQUILIBRIUM[9]

Equilibrium and its absence, disequilibrium, are concepts familiar in some degree to all students, from their study of economics or of other social or physical sciences. The definition of equilibrium in the physical sciences as a state of balance between opposing forces or actions applies without modification in the field of economic theory. Disequilibrium in turn simply be-

[8]Among other things, this illustration glosses over the question of how investment is distributed by industry. If the increase should be concentrated in a few strategic industries, aggregate investment in the next year may be adversely affected by a small increase in the aggregate stock of capital this year. This might be the case if these industries discovered that they had overexpanded facilities relative to final demand for their products.

[9]For more on the meaning of equilibrium, see C. A. Tisdell, *Microeconomics—The Theory of Economic Allocation*, Wiley, 1972, Chapter 4. See also F. Machlup, "Equilibrium and Disequilibrium: Misplaced Concreteness and Disguised Politics," in *Economic Journal*, March 1958, pp. 1–24, reprinted in the author's *Essays in Economic Semantics*, Norton, 1967.

comes the absence of a state of balance—a state in which opposing forces produce imbalance.

Since in economics we are continuously dealing with variables whose values change over time, the state of balance that defines equilibrium may perhaps be better expressed as a state of no change over time. This is not to say that economic equilibrium is a state of absolute rest, a motionless state in which no action takes place; rather, it is a state in which there is action, but action of a repetitive nature. Each time period exactly duplicates the preceding time period. This state of equilibrium is maintained, even though the forces acting on the system are in a continuous state of change, as long as the net effect of these changing forces is such as not to disturb the established position of equilibrium.

Let us turn for a moment to microeconomic theory and consider the ordinary supply-and-demand analysis of price determination for a single commodity in which quantity supplied varies directly with price and quantity demanded varies inversely with price. In Figure 3-1, supply, S, and demand, D, are in equilibrium only at a price of OP and a quantity of OA. At any price higher or lower than OP, there is disequilibrium: At any price above OP there will be an excess of quantity supplied over quantity demanded, and at any price below OP an excess of quantity demanded over quantity supplied. In this particular model, in the event of disequilibrium, the forces are such as to move price back to the equilbrium level of OP and quantity back to the equilibrium level of OA.[10]

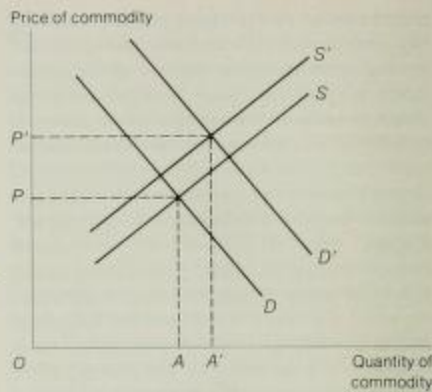


FIGURE 3-1
Supply of and demand for a commodity

Supply and demand are functions that indicate the different quantities of a commodity that will be supplied and demanded at various prices *for a particular time period.* As flow variables, supply and demand may be expressed in terms of quantity per minute, hour, day, week, or any other time period. If supply and demand in each time period are the same as in the preceding time period, the equilibrium quantity of the commodity purchased or sold will be OA and the equilibrium price will be OP, one time period after the other. The market is in balance, but it is not motionless, for sellers are continually bringing more of the commodity to market and buyers are continually taking more of it away. In other words, the market is in equilibrium; there is no change in the magnitude of the price and quantity variables.

Over time, of course, changes in supply and demand take place. Depending on the direction and the magnitude of the changes in supply or demand or both, equilibrium price and quantity may increase or decrease, with price and quantity changing in opposite directions or in the same direction. S' and D' in Figure 3-1 illustrate this last possibility. The new equilibrium price becomes OP', and the new equilibrium quantity

[10]Throughout this discussion it is assumed for simplicity that the particular equilibrium indicated by any pair of supply and demand curves will be attained as long as those curves remain unchanged for whatever time period is required for the adjustment process to work itself out. Actually, the mere fact that such an equilibrium exists does not necessarily mean that the system will move to it even over time. What happens depends on the nature of the dynamic process by which the system adjusts to a disequilibrium, and this process is not necessarily one that carries the system to the equilibrium position. For an introduction to this complex subject, see W. J. Baumol, *Economic Dynamics*, 3rd ed., Macmillan, 1970, Ch. 7.

becomes OA'. As long as a supply curve sloping upward to the right intersects a demand curve sloping downward to the right, any possible change in supply and demand will define a new equilibrium price and a new equilibrium quantity at the point of intersection of the two curves.

In practice, the new equilibrium price and quantity are not instantaneously established. The process takes time, and during this time price and quantity are changing, and the market is by definition in disequilibrium. If the changes in supply and demand are frequent, sizable, or erratic, equilibrium may never be established. Before the market can reach that price-quantity combination that represents equilibrium for one set of supply and demand conditions, the supply and demand conditions change. In such a situation, the market is forever moving toward equilibrium, but equilibrium has become a shifting, evasive goal that always recedes before it can be reached. However, even for markets like this that are in continuous disequilibrium, the concept of equilibrium is a valuable tool of analysis. If at any point in time an equilibrium position exists, this at least tells us which way the system is going to move next, even though we know that before the system gets to the equilibrium position toward which it is momentarily headed, it will be detoured by a change in the forces that change the equilibrium position.

Figure 3-1 was chosen to illustrate the concept of equilibrium because it is the simplest possible microeconomic model of a system with an equilibrium solution. This model contains only three variables—quantity of the commodity supplied, quantity of the commodity demanded, and price of the commodity—and only three relationships among these variables. Two are functional relationships: Quantity demanded is an inverse function of price, and quantity supplied is a direct function of price. The third relationship specifies the condition necessary for equilibrium: The quantity that suppliers wish to sell must be equal to the quantity that demanders wish to purchase, or, in brief, supply must equal demand. All the variables that bring about

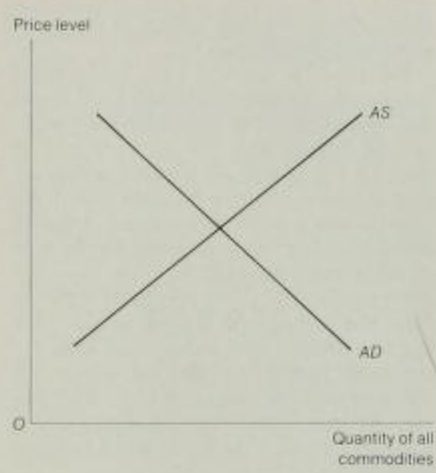


FIGURE 3-2
Aggregate supply and aggregate demand

shifts in the supply and demand curves, such as buyers' incomes, their tastes, the prices of other commodities, and the prices of inputs used in producing the commodity, are assumed to remain temporarily unchanged in order to focus attention on the way in which the equilibrium price is determined under given conditions of supply and demand.

Although the model for a single commodity is the more familiar, there is a macroeconomic model that parallels the microeconomic one. In the macroeconomic model, the coverage is not just of one of the many thousands of different goods and services supplied and demanded in markets but instead is of all of these combined. Therefore, in the macroeconomic model of Figure 3-2, the amounts measured along the horizontal axis are different aggregate quantities of goods and services. As each of the many goods and services in such an aggregate has its own price, what is measured along the vertical axis must correspondingly be the price level or an appropriately weighted average of the prices of all the goods and ser-

Jerk off

vices whose combined quantity is measured along the horizontal axis.[11]

The curves in Figure 3-2 are appropriately designated aggregate supply (*AS*) and aggregate demand (*AD*) to distinguish their much broader content from that of the curves in Figure 3-1. The intersection between the aggregate curves indicates the price level or the average price at which there is equality between the aggregate quantity of goods and services supplied and demanded. As in the case of a single commodity, at any price level higher or lower than that given by this intersection, there is disequilibrium. At a higher price level, the aggregate quantity supplied will exceed the aggregate quantity demanded and the price level will tend to fall to achieve equilibrium; at a lower price level, the opposite will be found and the price level will tend to rise to achieve equilibrium.[12]

If the price level is higher or lower than that required for equilibrium, it is evident that the prices of some individual goods and services will have to change in the process of attaining aggregate equilibrium. What is not so evident is that the prices of some individual goods and services may also change while the equilibrium price level and quantity remain unchanged at the levels indicated by the intersection of the given aggregate supply and demand curves. That is, shifts may occur in the individual supply and demand curves, and therefore changes may occur in the equilibrium price, equilibrium quantity, or in both for individual goods and services without the occurrence of any shift in the aggregate supply or aggregate demand curve. In this case the shifts in the curves for

[11] The step from the quantity of a single good or service to the quantity of all goods and services is a giant one that raises difficult questions. To merely illustrate with a brief look at one of these, consider the problem of measuring the change in the aggregate quantity of goods and services. One need go no further than two goods, say bread and cars, to see the nature of the problem. If there were only one good, bread, an increase from 100 to 200 loaves would clearly indicate a 100 percent increase in quantity. But take the case of two goods: the quantity of cars increases from 10 to 12 units and the quantity of bread increases from 100 to 200 loaves. Because both goods have increased in quantity, it is evident that aggregate quantity has increased, but one can not say specifically how large an increase has occurred unless one first assigns weights to each good to reflect the relative importance of each. This relative importance is best indicated by the prices at which the two goods sell in the market. If the price of a loaf of bread were \$1 and the price of a car were \$1,000 both before and after the change in quantities, weighting the original quantities by these prices and adding the two weighted quantities gives us (100 loaves × \$1) + (10 cars × \$1,000) = \$10,100 and weighting the changed quantities by the prices gives us (200 loaves × \$1) + (12 cars × \$1,000) = \$12,200. As \$12,200 equals 120.8 percent of \$10,100, the increase in aggregate quantity is 20.8 percent. If we wanted to show the two aggregate quantities in a diagram with aggregate quantity measured along the horizontal axis, each aggregate quantity would have to be expressed in dollars (of constant purchasing power as is assumed in our illustration). The two points along the axis would be \$10,100 and \$12,200 in the present case, and the difference between them would be one measure of the difference between the two aggregates of cars and bread. In the same way, the change in aggregate quantity for any other combination of quantities, including a decrease in both or an increase in one and a decrease in the other, may be calculated and the point for each located along the axis of a diagram.

As shown above, the measurement of the change in aggregate quantity presents no conceptual difficulty as long as the prices of the goods remain unchanged (or as long as all prices change by the same percentage), but this is not what is found in practice. With prices of different goods changing at different rates, the measurement of the change in aggregate output is not the unambiguous matter of our simple illustration. For a discussion of the so-called "index number problem" here confronted, see G. Ackley, *Macroeconomic Theory*, Macmillan, 1961, pp. 78–88, and, in a more technical vein, W. J. Baumol, *Economic Theory and Operations Analysis*, 4th ed., Prentice-Hall, 1977, pp. 350–53.

[12] The supply and demand curves for both the "one commodity" model of Figure 3-1 and the "all commodities" model of Figure 3-2 have been drawn here in a way intended to convey no more than the general notion that quantity supplied varies directly with price (one commodity) or with the price level (all commodities) and that the quantity demanded varies inversely with price (one commodity) or with the price level (all commodities). However, on certain assumptions the aggregate supply curve will show no upward slope (over the range up to the capacity level of output) and the aggregate demand curve will show no downward slope (i.e., the aggregate supply curve will be horizontal or perfectly elastic over the indicated range and the aggregate demand curve will be vertical or perfectly inelastic). The construction of the simple model of output determination in Part Two is based on assumptions that result in aggregate supply and demand curves of this kind. In Part Four other assumptions are adopted which result in curves more like those in Figure 3-2.

particular goods are such that the increase in the equilibrium quantity of some goods is just matched by the decrease in the equilibrium quantity of others, and the rise in the equilibrium price for some goods is just matched by the decrease in the equilibrium price for others. In other words, there may be shifts in the supply and demand curves for individual commodities that are offsetting and thus leave the aggregate supply and aggregate demand curves unchanged. There is, however, no need for the changes to be offsetting in this way, and shifts in the demand curves for some individual items may on balance produce some shift in the aggregate demand curve and similarly for shifts in the supply curves of some individual items. Therefore, the combination of price level and aggregate quantity which identifies an initial macroeconomic equilibrium can be displaced by shifts in supply and demand curves for some individual goods and services.

The idea of macroeconomic equilibrium may also be illustrated in a different way, through the concepts of stocks and flows examined in the previous section. Suppose water flowed into a reservoir at a rate of 100,000 gallons per day and out of the reservoir at a rate of 90,000 gallons per day. These flows would be described as equilibrium flows as long as they did not vary in size from day to day or over the period of time considered relevant. This produces *flow* equilibrium, but it necessarily also produces a disequilibrium in the *stock* of water. If the stock of water were measured at the same point in time each day, the gauge would show that the stock was growing by 10,000 gallons each day. Since the stock is changing, there is stock disequilibrium; since the flows are constant, there is flow equilibrium. Stock disequilibrium is thus logically consistent with flow equilibrium. Over time, however, a sufficient change in stock will begin to affect the previously constant flows. Unless the stock of water is to be permitted to overflow the banks of the reservoir, there must be a change either in the inflow (from 100,000 to 90,000 gallons per day) or in the outflow (from 90,000 to 100,000 gallons per day) or in both (to 95,000 gallons per day). If changes of this sort are made in the size of the flows, the system will be one in which both flows and stocks are in equilibrium.

An analogous situation is found in the flow of investment (capital goods produced), the flow of capital goods consumed, and the stock of capital goods. Gross investment at a constant rate of \$95 billion per year and capital consumption at a constant rate of \$55 billion per year define a flow equilibrium. These flows also define a stock disequilibrium in which the stock of capital increases every year by the amount of \$40 billion.[13] This is one indication that this is a "growing" economy if we measure economic "growth" by the accumulation of capital. In contrast, an economy exhibiting equilibrium in both flows and stock, with, say, gross investment of \$55 billion and capital consumption of \$55 billion per year, is a "stationary" economy if we define a "stationary" economy as one whose stock of capital neither increases nor decreases over time.

Flow equilibrium may thus be described as short-run equilibrium, and both flow and stock equilibrium may be described as long-run equilibrium. Since stock equilibrium cannot exist without flow equilibrium, long-run equilibrium cannot exist without short-run equilibrium. In short-run equilibrium, we disregard the disequilibrating effects that flows produce on stocks and consider only the conditions necessary to achieve flow equilibrium. In long-run equilibrium, however, the countereffects produced on flows by disequilibrium in stocks must be recognized, and conditions for full equilib-

[13]This conclusion of a stock disequilibrium follows from the definition of disequilibrium as an *absolute* change in the variable. If investment, capital consumption, and the stock of capital grow at such rates that the *ratio* of stock of capital to the flow variables does not change period by period, then, although the stock of capital is changing in absolute terms, it is not changing relative to flows. From this emerges a different and more complex definition of equilibrium as constancy in the ratio of capital stock to the relevant flows. Under this definition, then, what is disequilibrium in absolute terms may be equilibrium in relative terms. In this book, equilibrium is a position of no change in absolute terms unless otherwise noted.

rium encompass those necessary for both flow and stock equilibrium.

An economic theory or model abstracts from the infinite complexity of the real world by establishing what are believed to be the significant relationships among a limited number of variables deemed relevant to the problem at hand. The concept of equilibrium is a valuable tool of theory because it identifies a position in which the values of the model's variables are in balance. This helps simplify the complexity of the real world, where these same variables may actually be in continuous short- and long-run disequilibrium. Disequilibrium is also a valuable tool of theory but in a different sense, for by simplifying less, it more closely approximates economic reality. In fact, it may be said that short-run equilibrium analysis is a maximum in simplification and long-run disequilibrium analysis is a minimum in simplification. The more difficult branch of macroeconomic theory is therefore that which deals with systems in long-run disequilibrium by admitting into the analysis continuing changes in both flows and stocks.

At the end of the previous section we indicated that we would be primarily concerned with elementary macroeconomic theory in which changes in flows are considered but changes in stocks are not. For the same reason, the models we will consider will for the most part be those with an equilibrium solution. In other words, we will confine ourselves largely to short-run equilibrium models.

STATICS AND DYNAMICS[14]

We have noted that stocks and flows are the two types of variables found in economic models and that equilibrium and disequilibrium are the two possible positions of such models at any point in time. The actual position at any point in time is determined by the values attached to the variables that are parts of the model. Now let us examine briefly the two general methods employed in the construction and analysis of economic models—statics and dynamics.

These terms have been defined in somewhat different ways by different economists. One definition that conveys the meaning of dynamics in nontechnical language is offered by Professor Baumol: "Economic dynamics is the study of economic phenomena in relation to preceding and succeeding events."[15] Other definitions could be offered, but each in one way or another defines the essence of dynamics as the explicit recognition of time in the process of economic change.

In constructing formal models, one way of explicitly incorporating time is to split it up into periods and to examine how what happens in one period is related to what happened in preceding periods and to what is expected to happen in succeeding periods. In other words, the variables in dynamic models are said to be "dated." In contrast, the variables in static models all pertain to the same period of time, and there is no need to bother with dating. By dating the variables in dynamic models, we can investigate such things as how the amount of goods that businesspersons plan to purchase for inventory in a period may depend on the

[14]For an introduction to the concepts of statics and dynamics with special reference to Keynes's *General Theory*, see A. H. Hansen, *A Guide to Keynes*, McGraw-Hill, 1953, pp. 44–54. See also J. R. Hicks, *Value and Capital*, Oxford Univ. Press, 1939, Ch. 9; P. A. Samuelson, *Foundations of Economic Analysis*, Harvard Univ. Press, 1947, pp. 311–17; and F. Machlup, "Statics and Dynamics: Kaleidoscopic Words," in *Southern Economic Journal*, Oct. 1959, pp. 91–110, reprinted in the author's *Essays in Economic Semantics*, Norton, 1967.

[15]W. J. Baumol, *Economic Dynamics*, 3rd ed., Macmillan, 1970, p. 4.

amount of their sales in a previous period or on the amount of change in their sales between two previous periods. In turn, we can also investigate to what degree the sales volume in a previous period, or the change in sales between two periods, is influenced by the level of income of the economy in that previous period or by the change in the level of income between periods. In short, through this technique dynamic analysis is able to trace the changes in the values of the variables *over time*. The change in each variable from one period to the next is determined in a specified way by changes in the other variables included in the model.

Since statics ignores the passage of time, it is powerless to explain the *process of change* in a model. It can indicate the position of the model for a given period, but it cannot, except in a special case, tell exactly what the position will be in any other period. It is in the special case where the model is not changing at all but is simply repeating the same motion period after period that static analysis can reveal both where the system is in the present period and precisely where it will be in any future period—namely, exactly where it is in the present period. This special case is termed "stationary equilibrium," because the equilibrium position remains unchanged from one period to the next.

Pure static analysis is applicable only to a model in which a single, unshifting equilibrium position is established by the relationships among the variables. Applying statics to such a model in a period when it is in disequilibrium can only tell us for that particular period the values of the variables that are changing from that period to the next. Statics can explain *why* this is a disequilibrium, what relationship among the variables is necessary for equilibrium, and in what direction the system will next move. Given the fact that a single, unshifting equilibrium position exists, it may describe in general terms where the system must move to reach this predetermined equilibrium position. Statics cannot explain, however, the *actual process*, step by step or period by period, that the system follows over time in getting to that equilibrium position.

As the static method of analysis applies to models in equilibrium, the dynamic method of analysis applies to models in disequilibrium. Dynamics traces the process of change in the values of a model's variables over time, and a system in disequilibrium is, by definition, one whose variables are changing in value. Hence, to analyze a model in disequilibrium, we must use dynamics, the method that is capable of following the system from one point of disequilibrium to another toward an eventual equilibrium position or through an unending succession of disequilibrium positions.

To illustrate, let us return again to the microeconomic supply-and-demand model of Figure 3-1 discussed in the previous section. If for a given period the price–quantity combination is other than the equilibrium combination, price and quantity must change. Since we have assumed that there is an equilibrium position, the changes over time will be changes that are working toward this equilibrium price–quantity combination. Given the original supply and demand curves and assuming that the very process of working toward the indicated equilibrium of supply and demand will not cause a shift in either the supply or demand curves and therefore in the equilibrium position, static analysis can identify the equilibrium position and describe in general terms how the system will move to this position. If we were given more information about the way this market operates —much more than just the market's original supply and demand curves—dynamic analysis could be used to do what statics cannot. Dynamics could trace, period by period, the changes in the values of the variables as they moved through successive disequilibrium positions toward the single price–quantity equilibrium position.[14]

[14]As noted in footnote 13, we have assumed for simplicity that the equilibrium price (*OP*) and quantity (*OA*) will be reached. However, it is possible that in any period pur-

Comparative Statics

We have noted that the static method is meaningful only when applied to models with equilibrium positions. We also know that the economic forces that determine the equilibrium position for a model may be expected to change over time so as to displace the original equilibrium and, under certain conditions, to lead to the establishment of a new equilibrium. Given an initial position of equilibrium and some specified changes in underlying forces, if it is possible to determine how these forces affect the position of the new equilibrium, one can compare the two equilibrium positions and explain the change between the two in terms of the changes in forces. It is the analysis of this particular kind of change, from one equilibrium position to another, that may be handled by the method of comparative statics.

Consider once again the supply-and-demand model of Figure 3-1. The original equilibrium is defined by the intersection of the supply curve, *S*, and the demand curve, *D*. Suppose that changes in conditions external to the model, such as changes in income, buyers' tastes, prices of competing products, or prices of inputs used in production, cause the supply curve to shift to *S'* and the demand curve to shift to *D'*. Through the method of comparative statics, we can show the direction and the magnitude of the change in equilibrium price and quantity that follows from changes in the underlying forces that cause the shifts in the supply and demand curves. In our example, the changes in these forces are such as to raise equilibrium price from *OP* to *OP'* and equilibrium quantity from *OA* to *OA'*. Comparative statics can also tell us the magnitude and the direction of change in equilibrium price and quantity if the only shift that occurs is in the demand curve or in the supply curve. For example, if in our diagram a rise in consumer income were to shift the demand curve upward from *D* to *D'*, we could confidently predict that, with the supply curve, *S*, as shown, equilibrium price must rise. Thus, we have the comparative statics result that, with an upward-sloping supply curve, a rise in income that shifts the demand curve upward must raise equilibrium price.

We can conduct the same sort of analysis for changes in supply or demand other than those illustrated in Figure 3-1. However, comparative statics is adequate for this task only when in each case a new equilibrium position succeeds the old. Comparative statics is inadequate for the task when, as a result of changes in underlying economic forces, a system goes into a state of continuous disequilibrium. Furthermore, even if a new equilibrium does succeed the old, comparative statics is incapable of explaining the path followed by the system over time in getting from the old position of equilibrium to the new. In other words, comparative statics bridges the gap between equilibrium positions in one instantaneous jump, but it tells us nothing about *how* we got from one position to the other. In reality, since positions of disequilibrium are more the norm than the exception, we are likely to be more interested in the path followed between positions of equilibrium than in the positions themselves, and only dynamic analysis can handle this task.

To summarize the relationships between statics and dynamics and the concepts examined

---

chases and sales made at prices other than the equilibrium price will cause shifts in the supply and demand curves in the next period. This means that transactions in the present period at disequilibrium prices can in the next period produce a change in the equilibrium price. Statics is forced to circumvent this problem in some way. One way is to assume that the original equilibrium price is "instantaneously" reached; another is to assume that all purchases and sales are tentative rather than final until the particular price at which there is equilibrium is arrived at by all buyers and sellers. The latter process is sometimes referred to as "recontract." Without such unrealistic assumptions, which statics is forced to make, purchases and sales made at disequilibrium prices may set into motion a process that never reaches equilibrium, one disequilibrium price succeeding another. This may occur, for example, if a change in price in the present period gives rise to expectations among both buyers and sellers that there will be further price changes in the same direction during following periods. An analysis of a market such as this is possible only with the methods of dynamics.

earlier in this chapter: The variables found in an economic model are either stocks or flows. Any given model may include only flow variables or both flow and stock variables. Certain relationships are postulated among these variables, such that the value of one variable is a function of the value of one or more of the other variables in the model. If all variables are considered in the same time period, the relationships are all *static;* if they cover different time periods, the relationships are *dynamic*. Thus, in the supply-and-demand models as presented, the relationships are static; all the variables in the models are considered during the same time period.

Any given set of relationships may or may not produce an equilibrium solution for a model. In the microeconomic supply-and-demand model, equilibrium requires that quantity of the commodity supplied be equal to quantity of the commodity demanded. If there is a pair of values for these two variables that will equate the two, the price at which they are equal is the equilibrium price. If there is no such pair of values, there is no equilibrium price. The resultant model is a disequilibrium model in which the relationships between quantity supplied and quantity demanded must lead to constantly changing prices and quantities as transactions are carried through period after period at prices other than an equilibrium price.

If the model has an equilibrium solution, it may be analyzed by the static method. If it has no equilibrium solution, it can only be analyzed by the dynamic method. If the model is such that one equilibrium position, if upset by a change in some variable, will tend to be succeeded by a different equilibrium position in a manner that can be calculated from the change in the disturbing variable, the change from one equilibrium position to the next may be analyzed by the method of comparative statics but the actual path followed between equilibrium positions may not.

part two

The Simple Keynesian Model of Income Determination

chapter 20

4 Consumption and Investment Spending

A basic proposition of Keynesian theory is that the equilibrium level of income and output depends on the economy's aggregate demand for output. If aggregate demand is not sufficient to call forth the level of output that requires the employment of all available workers for its production, unemployment results, and production of goods and services is below its potential. If aggregate demand is just sufficient, full employment results, and production is at full potential. If aggregate demand is excessive, inflation results as well as full employment, and production is at full potential. However, any level of output, ranging from that which calls for full employment of the labor force to that which imposes idleness on a large part of the labor force, is a possible equilibrium level. Given this wide range of possible equilibrium levels, the actual equilibrium level in any time period is determined by the aggregate demand for that period.

The conclusion that aggregate demand determines the level of output follows from the completely passive role assigned to aggregate supply in the simple Keynesian theory. As shown by the aggregate supply curve, *AS*, in Figure 4-1, that theory assumes that aggregate supply is perfectly elastic up to the full employment level of output, Y_f, and is then perfectly inelastic at that level. Firms, in the aggregate, will supply various levels of total output up to the level Y_f at a constant price level, here shown as P_1. With an aggregate supply curve like this, it can only be shifts in the aggregate demand curve, such as from AD_1 to AD_3, or AD_4 to AD_2 or any other such shifts, that bring about changes in the level of output. At the same time, shifts in the aggregate demand curve over the range of output up to the full employment level cannot bring about changes in the price level. For levels of output below the full employment level, changes in the price level can come about only through shifts in the aggregate supply curve.

In this part of the book we shall assume that the aggregate supply curve does not shift. Therefore, shifts in the aggregate demand curve are movements along an unshifting aggregate supply curve, and as long as they are restricted to the perfectly elastic portion of the supply curve, they affect only the output level and not the price level. However, once the full employment level of output is reached, a further rightward shift in the aggregate demand curve will no longer result in additional output but will instead only pull up the price level. The infla-

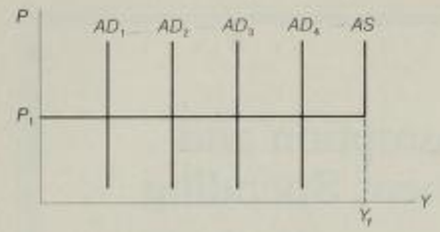


FIGURE 4-1
Aggregate supply and aggregate demand curves

tionary process set into motion by excess aggregate demand and what it leads to will be examined in Chapter 23. Here we merely state without explanation that the price level stability obtained along the horizontal portion of the unshifting aggregate supply curve is replaced by inflation once the aggregate demand curve pushes the economy beyond that range.

The elementary Keynesian model simplifies in one way by assuming a perfectly elastic aggregate supply curve over the range indicated, and in another way by assuming a perfectly inelastic aggregate demand curve. A perfectly elastic supply curve means that the aggregate output supplied by firms does not vary with the price level, and a perfectly inelastic demand curve means that the aggregate output demanded by buyers does not vary with the price level. In introducing the terms, aggregate demand and aggregate supply, in connection with the concept of equilibrium in Chapter 2, we used an upward sloping aggregate supply curve and a downward sloping aggregate demand curve. These are the more familiar shapes of these curves which will be derived and employed in Part Four. However, for the simple Keynesian model to be developed in this part of the book, we simplify in the way noted.

This is not to say that the simplifications adopted here, namely, an aggregate supply curve perfectly elastic over some range and an aggregate demand curve perfectly inelastic throughout, make this model one only distantly related to reality. However, offhand that might appear to be the case, especially because one expects demand curves in the real world to be less than perfectly inelastic. Every student learns very early the "law of demand," which states that the market demand curve for a particular good, say A, slopes downward to the right or, in terms of elasticity, that demand is not perfectly inelastic. He is therefore likely to draw the same conclusion for the aggregate demand curve. But to do so is to overlook something else learned early: the fallacy of composition. What is true for the demand curve for any one good or service is not necessarily true for the demand curve for the aggregate of goods and services.

The market demand curve for *A* or any other single good is drawn on the assumption that other things remain unchanged, including the incomes of the persons in the market and the prices of other goods in the market. On this basis it is easy to show that a lower price for a typical good, *A*, will be accompanied by an increase in the quantity of *A* demanded, because persons in the market will buy more of this good as they substitute it for other goods whose prices are unchanged. By the same reasoning, a rise in the price of *A* will be accompanied by a decrease in the quantity of *A* demanded. As most goods have more or less close substitutes, the demand curves for different goods typically slope downward to the right.

The same line of argument does not apply in the case of an aggregate demand curve. It is not as unreasonable as it may at first seem for such a demand curve *not* to show the downward slope that a market demand curve for a particular good or service shows. Suppose the price level falls by 10 percent (and for simplicity assume also that the price of every single good falls by this same percentage). There is a vast difference between such a fall in the price level and, say, a 10 percent fall in the price of one good (which makes up a negligible part of the

total of all goods and services) in terms of their effect on aggregate income. While it is reasonable to disregard the effect on aggregate income of the fall in the price of a single good, this is not true in the case of a fall in the price level. If the price of *A* falls by 10 percent, the sale of any specific number of units of *A* will generate 10 percent less income in the form of wages, salaries, profits and other factor incomes in the industry producing it than would the same number sold at the higher price. But because the workers, managers, and owners of the firms producing *A* account for a negligible part of the demand for *A*, the fact that, for any particular quantity of the good, less income is earned by them at a lower price for the good has virtually no effect on the number of units of that good demanded at that price. Therefore, a lower price will, via substitution as above described, be accompanied by an increase in the number of units demanded.

Consider now the case of a 10 percent decrease in the price level. Just as the income generated by the sale of, say, 100 units of *A* will fall by 10 percent if the price of *A* falls by 10 percent, the amount of income generated by the sale of, say, 100 million "units" of all goods will also fall by 10 percent if their price level falls by 10 percent. But unlike the case of a single good, we cannot assert that a 10 percent decrease in the price level will increase the aggregate quantity of goods demanded. Although such a result might at first glance be expected, a further look suggests that the same number of units may be demanded at the lower price level as was demanded at the higher price level. A larger quantity of one good will be demanded if its price falls because buyers will substitute this now lower priced good for others whose prices are unchanged, but substitution of one good for another means an increase in the quantity of one good demanded and a decrease in the quantity of another demanded. Here we are looking at the total of goods demanded and there is no increase in that total through substitution. Nor can we expect an increase in the total on the grounds that incomes will buy a larger total amount of goods at a lower price level for goods, because the total nominal income of the public is down by the same percentage as the price level. In real income terms, nothing has happened, and therefore there is no reason to expect a real change, such as a larger aggregate quantity of goods to be demanded at a lower price level. Looked at in this way, it is not altogether unreasonable to work with an aggregate demand curve that is perfectly inelastic with respect to the price level as in Figure 4-1.[1]

An aggregate demand curve like any of those in Figure 4-1 is the total demand for the many thousands of different kinds of output that a modern economy produces. An aggregate demand curve is also the sum of the demands by the different sectors in the economy. The traditional approach to the theory of income determination seeks to explain aggregate demand by explaining the spending of each of the major sectors into which the economy is divided: households, businesses, government, and the rest-of-the-world. To develop the simple Keynesian model in steps, this and the following chapter will be limited to an economy in which there are only households and businesses. Government and the rest-of-the-world are assumed not to exist for the time being. Aggregate demand is thus determined by consumption spending and private domestic investment spending. To explain what determines aggregate demand in any time period, we must first develop the essentials of the theory of consumption spending. This is done in the section that follows. Then, by at first simply assuming investment spending to be constant at some fixed dollar amount, we can proceed without

[1]This does not deny that there are other forces at work that give some elasticity to the aggregate demand curve. The extended Keynesian model that results from allowance for these forces will be developed in Part Four. Merely as an illustration of the relationships involved, an appendix to this chapter shows the way in which an elastic aggregate demand curve may be derived within the simple Keynesian model of this chapter.

further delay to find the level of aggregate demand. Because the elementary model is one in which the price level is determined by the aggregate supply curve and the output level is determined by the aggregate demand curve, determination of the aggregate demand curve directly yields the equilibrium level of output. In Chapter 5 we trace the process by which shifts in the aggregate demand curve occur in the two-sector economy and therefore how changes in aggregate output occur in our simple model of such an economy. Chapter 6 expands this model to three sectors by adding government, and Chapter 7 brings in the rest-of-the-world to provide the four-sector simple model.

Because the areas of consumption and investment spending receive minimum coverage in Part Two but are nonetheless the areas that present the major theoretical questions in the field of aggregate demand, Part Three is devoted to a detailed examination of some of the theoretical questions in the areas of consumption and investment spending.

CONSUMPTION SPENDING AND THE CONSUMPTION FUNCTION

What determines the aggregate amount of goods purchased by consumers in any time period? As we have seen, the simple Keynesian model does not include the price level as a determinant. A higher price level does not mean that less will be purchased nor does a lower price level mean that more will be purchased because it is assumed that the nominal income of households will rise or fall proportionally with a rise or fall in the price level. Therefore, households' real income in the aggregate is independent of the price level. In the elementary Keynesian model, it is the real income of households that basically provides the answer to the question posed above. A rise in real income will lead households to increase the amount of goods purchased and vice versa. This does not deny that there are many other less important determinants of real consumer spending, and some of these will be considered in Chapter 9. Here the assumption is that the aggregate amount of consumer goods purchased or the aggregate amount of real consumer spending is determined exclusively by the real income of consumers, i.e., by real disposable personal income.

Because the aggregate supply curve of Figure 4-1 gives us a stable price level of P_1 at all levels of output up to the full employment level, we do not have to differentiate between changes in nominal and real quantities as long as we limit attention to shifts in aggregate demand over the output range for which the price level is stable. For example, if AD_3 represent a 50 percent increase over AD_2, the product $P \cdot Y$ given by the intersection of AD_3 and AS will be 50 percent greater than the product of $P \cdot Y$ given by the intersection of AD_2 and AS, because AS is perfectly elastic over this range. This says that nominal income, the product of $P \cdot Y$, increases by 50 percent, but it also says that real income or Y also increases by 50 percent. Within this range, measuring the change in nominal income is the same as measuring the change in real income. Thus, in what follows, references to variations in income are not labeled real or nominal, the two being the same.

The Consumption Function

To consider how consumption expenditures are related to disposable income, we may begin by positing that consumption expenditures vary directly with disposable income. Second, we can be more specific and say something about how much such expenditures will vary as dis-

Higher Income — $\frac{S}{Income}$ — falls

posable income varies. Keynes did this in his "fundamental psychological law" which states that "men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income."[2] In other words, as income increases, consumers will spend part but not all of the increase, for they will choose to save some part of it. Thus, the total increase in income will be accounted for by the sum of the increase in consumption expenditures and the increase in personal saving, if we simplify by assuming that all consumer income goes into these two uses. Finally, can we be still more specific on the nature of this relationship? Although Keynes placed great confidence in the correctness of his "fundamental psychological law," he advanced with less confidence the argument that a smaller *proportion* of income will be consumed (or a larger *proportion* of income will be saved) as income increases. If this is true, it means not only that the absolute amount of saving will increase with increases in income as indicated by the "fundamental psychological law," but also that the ratio of saving to income will become greater with increases in income. Keynes felt that this was to be expected, *as a rule*, for, despite the fact that "the satisfaction of the immediate primary needs of a man and his family is usually a stronger motive than the motives toward accumulation," the latter "acquire effective sway when a margin of comfort has been established."[3]

It is this relationship between consumption and income advanced by Keynes that is employed in the simple theory of income determination to be developed here.[4] We will, in other words, proceed on the assumption that the absolute level of consumption varies directly with the level of income and that the fraction of income consumed varies inversely with the level of income.

Theoretical and Empirical Consumption Functions The relationship between consumption and income that emerges from these particular assumptions is referred to as a *theoretical consumption function*. As a tool of theory, a consumption function is somewhat analogous to an ordinary market-demand function for a single commodity, such as that in Figure 3-1. Just as a theoretical demand curve usually implies that the quantity of a commodity that will be purchased varies inversely with its price, all other things (including income) being unchanged, so the theoretical consumption function here employed holds that aggregate consumption varies directly but not proportionally with consumer *income*, all other things (including prices) being unchanged. The theoretical consumption function that we shall draw, like the theoretical demand function of Figure 3-1, is not derived from any actual statistical data. It is nothing more than an attempt to describe in general terms, on the basis of the assumptions stated above, a typical functional relationship between two variables, all other things being unchanged. Actually, for this purpose no explicit dollar amounts need be indicated; for example, they are not used in Figure 3-1.

A distinctly different type of consumption function, the historical or empirical consumption function, will be examined in Chapter 8. Here we may note that the simplest form of an empirical consumption function describes the statistics of income and consumption for each year over a period of years. Since these are recorded quantities, nothing can be assumed to have remained unchanged over the period of years involved. As a result, the actual level of consumption that accompanies the actual level of income for any year reflects every factor that

[2] John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936, p. 96.

[3] *Ibid.*, p. 97.

[4] It may be noted here that the particular relationship between consumption and income advanced by Keynes is perhaps the first statement of what later came to be known as the absolute income theory. This theory will be examined in Chapter 8 along with the relative income and permanent income theories.

influenced consumption expenditures during that year, not just the year's disposable income. In fact, the nonincome factors may be such as to cause the actual level of consumption in any one year to rise above that of the preceding year, despite a fall in the actual level of disposable income. On the basis of income alone, a decrease in consumption would be expected between these two years, but what otherwise would have been a decrease was more than offset by these other factors that made for a net increase. We will note actual cases like this later, when we look into empirical consumption functions. However, for our present purpose, namely to develop the simple theory of income determination, we need only the theoretical consumption function, which abstracts from all nonincome influences on consumption and posits a relationship between consumption and disposable income that satisfies certain assumptions.

The line labeled C in Part A of Figure 4-2 is one of many possible theoretical consumption functions that satisfy the specific assumptions stated above. The other line is a 45° guideline; any point on this line is equidistant from the vertical and horizontal axes. For example, point K on the 45° line represents a value of income, or $Y = 160$, on the horizontal axis and an equal value of consumption plus saving on the vertical axis.

Since by definition any portion of disposable income that is not consumed must be saved, then, given the consumption function, the guideline, and the hypothetical amounts laid off on the axes, we can tell at a glance how people plan to allocate any given level of disposable income between spending and saving. For example, at $Y = 160$ (point M), total income is given by $MK = 160$. Since consumption is $ML = 140$, saving must equal the balance of income, $MK - ML = LK$, or $160 - 140 = 20$.

The specific consumption function in Figure 4-2 is drawn on the assumption that there is some level of income at which planned consumption is exactly equal to income. This is referred to as the "break-even" level of income; here it occurs at level 80, since the consumption function cuts the guideline at this level. At any higher level of income, people collectively feel well enough off to save some part of their aggregate income. Above the break-even level, therefore, the consumption function lies below the guideline, and the vertical distance between the two lines equals the amount of saving for that level of income. At any level of income below 80, people collectively spend more than their aggregate income. In this situation, the consumption function lies above the guideline and the vertical distance between the two lines equals the amount of dissaving, or the excess of consumption over income, at that level of income. At an income level of 40, the consumption function is 10 above the guideline; the excess of consumption over income is thus 10 at this income level.

Average Propensity to Consume The average consumption–income relationship is defined by the ratio C/Y for different levels of Y. For the function in Figure 4-2, at Y of 40, we have C of 50, so that $C/Y = 50/40$, or 1.25. At Y of 80, we have C of 80, so that $C/Y = 80/80$, or 1, the break-even ratio. At Y of 160, we have C of 140, so that $C/Y = 140/160$, or 0.875. The C/Y ratio could be computed for any other level of Y in similar fashion. However, from what is given, it is apparent that the ratio of C to Y in this consumption function decreases steadily as income increases, and vice versa. In other words, C increases less than proportionally with increases in Y, and vice versa. The C/Y ratio, one of two basic ratios that may be derived from the consumption function, is known as the *average propensity to consume*, or the APC.

Marginal Propensity to Consume If we know the APC at all levels of disposable income, we know how each level of disposable income will be divided between consumption and saving. Suppose, however, we also want to know how any given *change* in the level of income will be

divided between a *change* in consumption and a *change* in saving. The APC will not give us the answer directly, but the slope of the consumption function will. To see this, take any two levels of disposable income and call the difference between them ΔY. Then determine the amount of consumption at each of these two levels and call the difference between these two amounts of consumption ΔC. For example, if we take Y of 200 and Y of 220, ΔY is 20. With Y

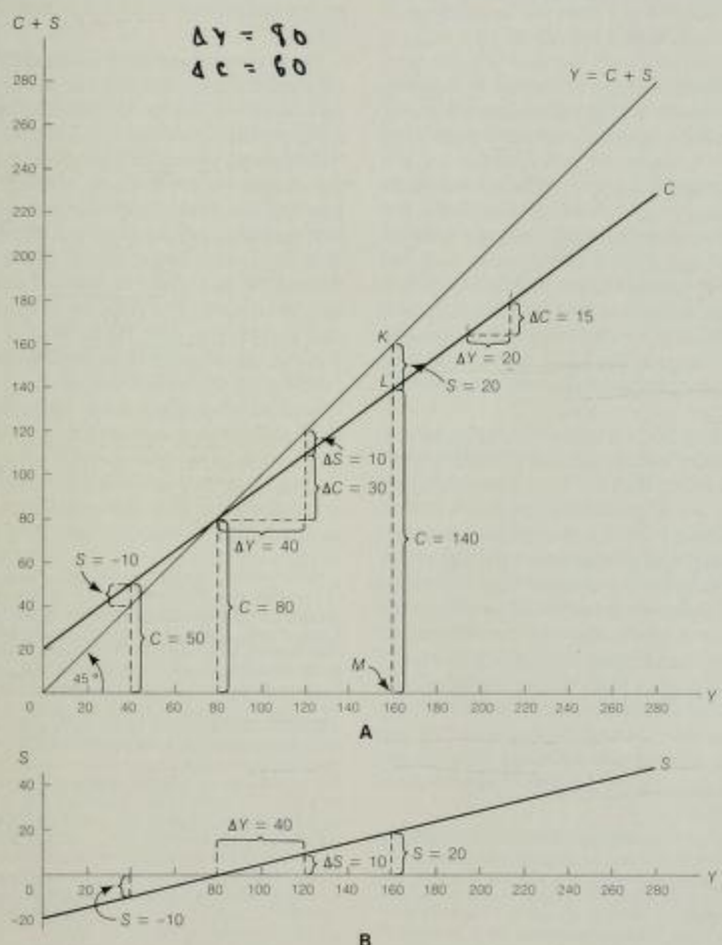


FIGURE 4-2
Consumption and saving functions

A.P.C. = $\frac{C}{Y}$
M.P.C. = $\frac{\Delta C}{\Delta Y}$

of 200, C is 170; with Y of 220, C is 185; therefore, ΔC is 15. Expressed as a ratio, $\Delta C/\Delta Y = 15/20 = 3/4$. In Figure 4-2, for any selected change in income taken anywhere along the income axis, the same result will be found; for every change of 4 in Y there will be a change of 3 in C, or a constant ratio of $\Delta C/\Delta Y = 3/4$. Geometrically, it should be clear that this ratio equals the slope of the consumption function. $\Delta C/\Delta Y$, the second basic ratio derived from the consumption function, is known as the *marginal propensity to consume*, or the MPC.

Note, however, that only if the consumption function is a straight line as in Figure 4-2 will the MPC be the same for any change in income. Any other straight-line consumption function with a slope different from that in Figure 4-2 will indicate an MPC larger or smaller than 3/4, but again an MPC that is the same for any change in income. The slope of the consumption function is the geometric representation of the MPC.[5]

Up to this point, we have not considered any evidence that would indicate whether a consumption function of the type given in Figure 4-2 is a realistic description of the way that persons divide their income between consumption and saving at different levels of disposable income. We will turn to this question in Chapter 8, but at the moment all we have is the hypothesis that the consumption–income relationship exhibits certain properties that may be summarized as follows. The MPC is positive but less than 1, this being Keynes's "fundamental psychological law." The MPC is the same for any change in income, this following from the assumption that the consumption function is a straight line. The APC is infinity at a zero level of income and declines steadily as income rises but is always greater than the MPC, this following from the previous assumptions plus the assumption that consumption remains positive no matter how low the level of income may fall.[6]

The Consumption Function—Equations Mathematically, the straight-line consumption function shown in Figure 4-2 may be described in terms of its intercept with the vertical axis and its slope with the aid of the simple equation of the straight line.[7] In the case of this consumption function, the intercept with the vertical axis indicates that C is 20 when Y is zero. For any level of income above zero, given the slope or the MPC as 3/4, C will be the 20 it would be at zero income *plus* 3/4 of the difference between zero income and any chosen level of income. This may be written in equation form as $C = 20 + \frac{3}{4}Y$. This theoretical consumption function may thus be thought of as the sum of two parts: an amount of consumption that is independent of the level of income, since it is the amount found even at zero income (20), and an amount of consumption that depends on the level of income, since it rises and falls by a constant

[5]If this is not apparent, keep the present straight-line consumption function anchored in its given position on the vertical axis and visually tilt it upward. The MPC will become greater than 3/4. Tilt it so that it becomes parallel with the guideline (which has a slope of 1) and the MPC becomes 1. Tilt it even more, and the MPC exceeds 1. Tilt the consumption function downward from its given position, and the MPC will become less than 3/4. Tilt it so that it becomes parallel with the horizontal axis, and the MPC becomes zero.

[6]Geometrically, the APC at any level of income is equal to the slope of a line from the origin to the point on the consumption function corresponding to that level of income. If such lines are drawn into Figure 7-1, it will be seen that the slopes of such lines start at infinity at zero income and fall steadily as income increases. Since the MPC equals the slope of the consumption function itself, any straight-line consumption function that cuts the vertical axis above the origin as in Figure 4-2 is a line whose constant slope is at all points less than the slope of a line from the origin to that point. In other words, for such a consumption function we have the property that $C/Y > \Delta C/\Delta Y$ at all levels of income. Note, however, that any straight-line consumption function that intersects the axes at the origin will be one for which $C/Y = \Delta C/\Delta Y$ at all levels of income. Finally, any straight-line consumption function that intersects the vertical axis below the origin will be one in which $C/Y < \Delta C/\Delta Y$ at all levels of income. For nonlinear consumption functions, the relationships between C/Y and $\Delta C/\Delta Y$ are more complex.

[7]The standard linear equation is $y = a + bx$, where b is the slope of its graph and a is its y intercept (the value of y at the point where $x = 0$; that is, the point at which the graph cuts the y axis).

A S/Y MPS = $\frac{\Delta}{\Delta Y}$

fraction (3/4) of any rise or fall in income. The first part is commonly described as autonomous consumption, C_a, and the second part as induced consumption. The subscript *a* is used here and elsewhere to designate a variable whose magnitude is autonomous or independent of the level of income.

This equation for the consumption function tells us everything that Figure 4-2 tells us. Just as we can find C for any level of Y by inspection in Figure 4-2, we can do the same by substituting any level of Y in the equation. Figure 4-2 shows that C is 80 when Y is 80, and the equation shows the same: $C = 20 + \frac{3}{4}(80) = 80$. Similarly, when Y is 160, the figure shows that C equals 140, and the equation shows the same: $C = 20 + \frac{3}{4}(160) = 140$. To find the APC or C/Y at any level of Y, we simply divide the original equation through by Y, or $C/Y = 20/Y + \frac{3}{4}Y/Y$, which is equal to $20/Y + 3/4$. When Y equals 80, $C/Y = 20/80 + 3/4 = 1$. When Y equals 160, $C/Y = 20/160 + 3/4 = 0.875$.

Since we will not restrict ourselves later to the consumption function with the intercept and slope of the function shown in Figure 4-2, a general equation for the linear consumption function may be given here:

$$C = C_a + cY$$

in which C_a is autonomous consumption or the amount of consumption when Y equals zero and the constant c is the slope of the function or the MPC. If we divide this equation through by Y, we derive the general equation for the APC.

$$\text{APC} = \frac{C}{Y} = \frac{C_a}{Y} + c$$

The hypothesis as to the way people divide their income between consumption and saving was summarized above as a set of properties. To satisfy these properties, it is required that C_a be positive and that c be positive but less than 1. This may be verified by noting that if C_a is negative, the proportion of income consumed increases as income increases; if c is greater than 1, the increase in consumption accompanying an increase in income exceeds the increase in income; and if c is negative, there is a decrease in consumption with an increase in income. All these possibilities conflict with the hypothesis.

The Saving Function

Part B of Figure 4-2 shows the saving function, which is the counterpart of the consumption function shown in Part A. In Part A the amount of saving at any level of income is the difference between the consumption function and the guideline. The saving function shown in Part B can therefore be directly derived from Part A.

When income is 80, we see in Part A that consumption is 80 and saving is 0; this is depicted in Part B by the intersection of the saving function with the horizontal axis at income of 80. When income is 40, consumption is 50, and saving is -10; the saving function lies 10 below the horizontal axis at income of 40. When income is 160, consumption is 140, and saving is 20; the saving function lies 20 above the horizontal axis at income of 160.

Average Propensity to Save The saving counterpart to the APC is the *average propensity to save*, or the APS. While the APC is the ratio of C to Y, the APS is the ratio of S to Y. Since Y itself is devoted to either C or S, it follows that the two ratios, C/Y and S/Y, must add up to 1. Thus when Y is 40, C/Y is 50/40, or 1.25, and S/Y is $-10/40$, or -0.25. Similarly, when Y is 160, C/Y is 140/160, or 0.875, and S/Y is 20/160, or 0.125.

Marginal Propensity to Save There is also a saving counterpart to the MPC. If, instead of looking at the ratio of S to Y at any level of Y, we look at the ratio of the change in S to the change in Y for any change in Y, we have what is termed the *marginal propensity to save*, or the MPS. Given a change in Y, ΔY, then $\Delta S/\Delta Y$ is the ratio of the change in S to the change in Y, just as $\Delta C/\Delta Y$ is the ratio of the change

in C to the change in Y. Since ΔY must be devoted to either ΔC or ΔS, the two ratios $\Delta C/\Delta Y$ and $\Delta S/\Delta Y$ must add up to 1.

If the MPC is positive but less than 1 and is the same for any change in income, then it follows by subtraction, since MPS = 1 − MPC, that the MPS must also be positive but less than 1 and that it must also be the same for any change in income. Furthermore if the APC decreases steadily as income rises, then the APS must increase steadily as income rises, because these two ratios also add up to 1 at all levels of income. Finally, if the APC is always greater than the MPC, it follows that the APS is always less than the MPS.[8]

The Saving Function—Equations As was the case for a straight-line consumption function, so a straight-line saving function can be described in terms of its vertical intercept and its slope with the aid of the equation of the straight line. The derivation of the equation for the saving function is analogous to that for the consumption function. When income is zero, saving is −20. For any level of income above zero, saving is −20 plus 1/4 of the difference between zero income and any chosen level of income. Therefore, $S = -20 + \frac{1}{4}Y$. To derive the equation for the APS, or S/Y, we simply divide through by Y, which gives us $S/Y = -20/Y + \frac{1}{4}Y/Y$, or $S/Y = -20/Y + 1/4$.

The general equation for the linear consumption function was given as $C = C_a + cY$, so the general equation for the linear saving function may be given as $S = S_a + sY$, in which S_a equals autonomous saving or the amount of saving at the theoretical zero level of income and s equals the marginal propensity to save.[9] If we divide this equation through by Y, we derive the general equation for the APS:

$$\text{APS} = \frac{S}{Y} = \frac{S_a}{Y} + s$$

The hypothesis about how people divide their incomes between consumption and saving was summarized above with a set of properties. To satisfy these properties, it is required in the equation for the saving function, $S = S_a + sY$, that the value of S_a be negative and the value of s be positive but less than 1. This may be verified by noting that if S_a is positive, the proportion of income saved decreases as income increases; if s is greater than 1, the increase in saving accompanying any increase in income exceeds the increase in income; and if s is negative, there is a decrease in saving with an increase in income. All these possibilities conflict with the hypothesis.

DETERMINATION OF THE EQUILIBRIUM LEVEL OF INCOME AND OUTPUT

The GNP identity for a two-sector economy was given in Chapter 2 as $C + S = \text{GNP} = C + I$ in which S includes business saving in the form of capital consumption allowances and I includes business investment spending before deduction of capital consumption allowances. If both of these are measured net of capital consumption allowances, the final product so measured is net national product. Since there is no government in this economy, national income

[8] Earlier, in describing the relationship between APC and MPC, we saw that for a consumption function of the type given in Figure 4-2, APC > MPC at all levels of income. Since APC + APS = 1 and MPC + MPS = 1, if APC > MPC at all levels of income, it follows that MPS > APS at all levels of income. For example, at $Y = 80$, APC = 1 and MPC = 3/4, while APS = 0 and MPS = 1/4. At Y of 160, APC = 7/8 and MPC = 3/4, while APS = 1/8 and MPS = 1/4.

[9] S_a equals $-C_a$, and s equals $1 - c$. This can be shown as follows: Since $Y = C + S$, $S = Y - C$. Substituting $C_a + cY$ for C, we have $S = Y - (C_a + cY)$. From this, $S = Y - C_a - cY$, or $S = -C_a + Y - cY$, or $S = -C_a + (1 - c)Y$. Hence, since we have written $S = S_a + sY$, it follows that $S_a = -C_a$, and $s = 1 - c$ as was to be demonstrated.

equals net national product.[10] If we further assume that all firms are noncorporate, there are no undistributed profits, and personal income equals national income.[11] Again, since there is no government, there can be no taxes, and all personal income becomes disposable personal income. In this economy, disposable personal income equals net national product; every dollar spent during the time period for either consumption or net investment produces a dollar of disposable personal income. Disposable personal income must be devoted either to personal consumption expenditures or to personal saving. Since disposable personal income equals net national product, personal saving (the amount of unconsumed disposable personal income) must then equal investment (the amount of unconsumed net national product).

If we measure the results for any time period in this economy, we have the following identities:

$$\text{Net National Product} = C + I$$
$$\text{Disposable Personal Income} = C + S$$

and

$$S = I$$

in which S and I are understood to be net amounts.

Since net national product and disposable personal income are identical in this economy in any time period, we may refer to them interchangeably, disposable personal income being identical with the value of output and the value of output being identical with disposable personal income. If we designate both by Y, we may write:

$$Y = C + I$$
$$Y = C + S$$

and, as before,

$$S = I$$

[10]Strictly, this also requires that business transfer payments be zero, an assumption we make here.

[11]In a two-sector economy in which all firms are noncorporate, personal income would exceed national income by the amount of interest paid by consumers. The easiest way to avoid the complications this factor would otherwise bring into the analysis is to assume that interest paid by consumers is zero.

These are the fundamental accounting identities with which we will work in the two-sector economy. Note that these are the same identities developed in Chapter 2 except for the substitution of net investment for gross investment and correspondingly net national product for gross national product. As identities, they are composed of the *realized* values for the variables for any time period. Thus, by our accounting definitions, realized saving is identical with realized investment. However, as we will see, realized investment may be greater or less than the amount of investment *planned* by business persons, if the amount of investment planned by business persons differs from the amount of saving planned by income recipients. In what follows we will use the terms *realized* and *planned* to make this distinction. Later in the chapter we will turn to a more detailed examination of realized and planned investment.

Equilibrium Income and Output

What determines the economy's consumption and investment expenditures? According to the assumption we have worked with so far, disposable income is the sole determinant of consumption expenditures. What investment expenditures will be depends on factors yet to be considered. However, in order to get started, suppose that these factors are such that business persons plan to spend a total of 20 (billion dollars) per time period for additions to plant and equipment and change in inventories. No matter how they may have arrived at these plans, all that is essential here is to assume for the time being that the plans are independent of the level of output. In other words, at all levels of output, planned investment expenditures are fixed at 20, or the investment function is simply $I = 20$.[12]

[12]Although we are assuming for the time being that investment expenditures are not functionally related to any other variable, it is still convenient here to use the term "investment function" to parallel the term "consumption function." In Chapter 11 we will examine several theories of investment spending and will employ "investment functions" that express these specific theories.

To derive a function or curve which will show aggregate expenditures or aggregate spending at each level of output, we must add together the consumption and investment functions. This is illustrated in Part A of Figure 4-3 by the curve labeled $C + I$. Here the consumption

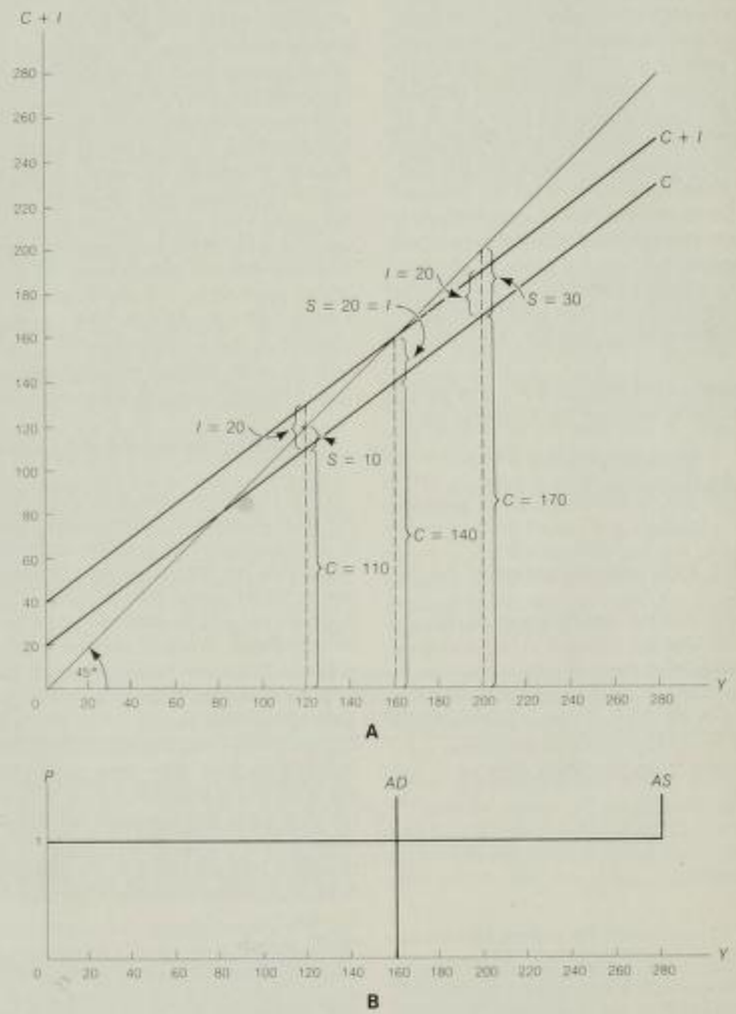


FIGURE 4-3
The equilibrium level of income and output: the $C + I$ approach

function is the same as that in Figure 4-2 which indicates the amount of planned consumption at each level of output. The investment function, which we have assumed is constant at 20 for all levels of output, is added to the consumption function. The resulting aggregate spending function is read as follows: If aggregate output and therefore aggregate real income were 120, aggregate spending would be 130 (110 + 20), or if aggregate output and therefore aggregate real income were 160, aggregate spending would be 160 (140 + 20).

It is important at this point to introduce a critical distinction whose understanding will prevent the confusion that commonly occurs in this area as a result of assigning the same name to two altogether different curves. The $C + I$ curve in Part A of Figure 4-3 shows the amount of expenditures or spending by consumers and businesses at each level of output. (In Chapters 6 and 7 we will introduce $C + I + G$ and $C + I + G + (X - M)$ curves which, respectively, show the amount of expenditures or spending on domestic goods and services by consumers, businesses, and government combined and by these three plus the rest-of-the-world combined.) Curves of this kind throughout this book will be referred to as aggregate spending or aggregate expenditure curves; they will not be referred to as aggregate demand curves as that term is reserved for an altogether different curve. The term "demand" in microeconomics refers to the schedule of the quantities of a particular commodity that will be purchased at various possible prices for that commodity. It not only sidesteps a source of confusion in macroeconomics but also maintains consistency between the use of the term in microeconomics and macroeconomics to reserve the term demand in macroeconomics to describe the schedule of the aggregate quantities of all commodities that will be purchased at various possible price levels for all commodities. It is this relationship which is shown by the aggregate demand curve in Part B of Figure 4-3. Thus, in the present model with its assumption of a stable price level, the aggregate spending or expenditure curve in Part A shows how the aggregate amount of goods purchased varies with the level of real income or output. The aggregate demand curve in Part B shows how the aggregate amount of goods purchased varies with the price level. Because the present model assumes that the aggregate amount purchased is perfectly inelastic with respect to the price level, the *AD* curve in Part B appears as a vertical line, and, for reasons to be explained, as a vertical line whose position is determined in Part A of Figure 4-3.

Given the simplifying assumptions adopted in the present model, the aggregate spending function in Part A of Figure 4-3 is all that is needed to determine the equilibrium level of aggregate output. Each dollar amount along the horizontal axis of both parts of Figure 4-3 indicates a different level of aggregate output. As shown by the aggregate supply curve in Part B, producers will supply at a price level of 1 an aggregate output of anywhere from zero to the upper limit of 280 at which the aggregate supply curve becomes perfectly inelastic.[13] Which of these possible levels of aggregate output will be the equilibrium level is determined in this model entirely in Part A as that particular level of output or Y at which $C + I$ is just equal to Y or at which aggregate spending is just equal to aggregate output and income. It will be seen that this occurs only at $Y = 160$. At a higher level of Y, $C + I < Y$, and at a lower level of Y, $C + I > Y$. With equilibrium established at 160 in Part A, it is appropriate to insert in Part B an

[13]The choice of 1 for the price level is arbitrary. It has the advantage of making the market value of any level of output equal to the figure given for that level of output on the horizontal axis of both parts of the figure. In contrast, if P were assumed to be 1.5 or 0.5 for example, the market value of any level of output shown on the horizontal axes would be respectively 1.5 or 0.5 of the figure given on the horizontal axis. However, the essential point to be seen in the present model is that the equilibrium level of output, which will be found to be 160, is the same whether the price level is 1, which makes the market value of that output equal to 160, or whether the price level is 1.5 or 0.5, which would make the market value of that same output equal to 240 or 80, respectively.

aggregate demand curve (perfectly inelastic in the present model) at $Y = 160$. The equilibrium level of output is 160 and the equilibrium price level is 1, for these are the levels given by the intersection of the two curves.[14]

To show why this is the case, let us look more closely at what the curves reveal. Suppose that business persons believe that during a given time period they can sell 160 in goods and that they produce this amount of goods during the period.[15] The aggregate supply curve shows that they were able to produce and supply up to 280 at the indicated price level of 1, but our assumption is that the actual amount produced is 160. With output of 160, disposable income will be 160; and with disposable income of 160, the consumption function in Figure 4-3 indicates that consumers will spend 140. Adding to this the 20 that business persons will spend, we have aggregate spending of 160 when output is 160. *Business persons produced aggregate output of 160 in the expectation that sales of output would total 160, and sales turned out to be exactly 160, so the plans of both sellers and buyers were realized.* Consumers with income of 160 purchase the 140 of consumer goods they plan to buy when income is 160; business persons purchase the 20 of investment goods they plan to buy. Both sectors purchase the amounts intended, and their purchases match exactly what business persons intended to sell. In short, the amount of goods purchased equals the amount supplied—which is one way of stating the condition for equilibrium in the level of output.

In Part A of Figure 4-3, the aggregate spending function, $C + I$, intersects the guideline at the equilibrium level of output. Aggregate output of 160, measured on the horizontal axis, is matched by an equal aggregate of spending 160 ($140 + 20$), on the vertical axis. Since any point on the guideline is equidistant from both axes, and since the condition for equilibrium in the level of output is that $C + I$ be equal to Y, it follows that the equilibrium level of output must be that level of output at which the aggregate spending function intersects the guideline.

Equilibrium may also be defined as that level of output at which planned saving equals planned investment. In Part A of Figure 4-4, planned investment is shown by an investment curve horizontal to the output axis and 20 above that axis to conform with the assumption that investment is 20 at all levels of output. The construction of the saving curve was explained earlier. It will be seen that planned saving is 20 at the output and income level of 160. Since planned investment is also 20, the amount that business persons choose to spend for investment goods exactly matches the amount of income of 160 that income receivers choose *not* to spend for consumer goods. Otherwise expressed, given that the dollar amount of income generated during a time period is equal to the dollar amount of goods and services produced during that time period, it follows that aggregate spending will have to be equal to that aggregate output of goods and services for the period if each dollar of that period's income that is saved or not spent on consumption is matched by a dollar spent on investment. If planned saving and planned investment are equal, aggregate spending and aggregate output are also equal.

[14]The present model, with the simplifications of a perfectly elastic aggregate supply curve (up to the full employment level of Y) and a perfectly inelastic aggregate demand curve, has all of its analytical content in Part A. All that Part B adds is the price level of output, and that is here arbitrarily set by whatever position is assumed for the aggregate supply curve. However, once this model is extended in one direction to incorporate an upward sloping aggregate supply curve and to derive rather than assume the position of that curve, and in another direction to derive a downward sloping aggregate demand curve, the part of the figure containing the aggregate supply and aggregate demand curves has analytical content. As will be seen in Part Four, it then becomes the part to which we must turn to find both the equilibrium level of output and the equilibrium price level. Although the part showing the aggregate supply and demand curves could be omitted from the present model, it is here presented to maintain consistency between the graphic framework for the present simple model and the model to be developed in Part Four.

[15]Not all goods, of course, are produced for sale to consumers and to other firms. Parts of some firms' production may be intended as additions to their own inventories. Such output is viewed as if it were sales of goods by firms to themselves. Thus, some part of estimated sales of 160 is made up of planned "sales" by firms to themselves.

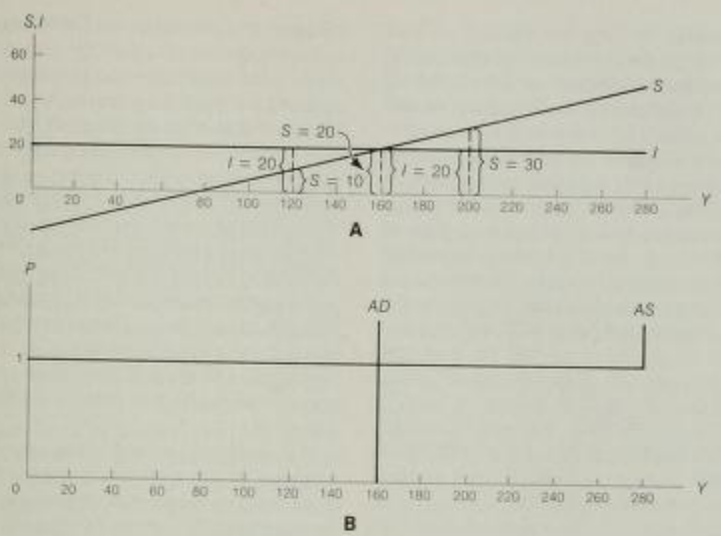


FIGURE 4-4
The equilibrium level of income and output: the S and I approach

As in Part B of Figure 4-3, an aggregate demand curve (perfectly inelastic) is erected in Part B of Figure 4-4 at the output level of 160. Graphically, we see that planned saving and planned investment can be equal only at the intersection of the two curves in Part A, and it is the intersection of these two curves, here occurring at an output of 160, which determines that the aggregate demand curve is positioned at 160 in Part B.

Equilibrium Income and Output—Equations

Taking as our condition for equilibrium the equality between $C + I$ and Y, we may also determine this equilibrium income and output by solving the equation $Y = C + I$, in which C and I refer to *planned* consumption and *planned* investment. Solution of this equation gives that level of output at which planned spending by persons on consumption out of the income earned in producing that output will, when added to the planned spending on investment, be just sufficient to purchase the amount of output actually produced. The amount income receivers plan to spend on consumption at any possible level of income is given for our economy by the equation $C = 20 + \frac{3}{4}Y$. The amount business persons plan to spend for investment is assumed to be fixed at 20 and is given by the equation $I = 20$. This gives us the following three equations:

$$Y = C + I \qquad [1]$$

$$C = 20 + \tfrac{3}{4}Y \qquad [2]$$

$$I = 20 \qquad [3]$$

Substituting [2] and [3] into [1] and solving for Y, we have

$$Y = 20 + \tfrac{3}{4}Y + 20$$

or

$$Y = 160$$

Alternatively, we find the equilibrium level of output to be that at which planned saving equals planned investment, or $S = I$. At this particular level of output, the amount *not* spent on *consumption* out of the income earned in producing that output will be exactly offset by an amount spent by business persons for *investment*. The amount income receivers plan to save at any possible level of income is given by the equation $S = -20 + \frac{1}{4}Y$. The amount business persons plan to spend for investment is again assumed to be fixed at 20; that is, $I = 20$. Thus, we have the following three equations:

$$S = I \qquad [1]$$

$$S = -20 + \tfrac{1}{4}Y \qquad [2]$$

$$I = 20 \qquad [3]$$

Substituting [2] and [3] into [1] and solving for Y, we have

$$-20 + \tfrac{1}{4}Y = 20$$

$$Y = 160$$

This confirms that only when Y equals 160 will aggregate spending equal aggregate output and planned saving equal planned investment. With C, I, and S all referring to planned amounts, we may summarize as follows:

Output:

$$Y = C + I$$

$$160 = 20 + \tfrac{3}{4}(160) + 20$$

Income:

$$Y = C + S$$

$$160 = 20 + \tfrac{3}{4}(160) + [-20 + \tfrac{1}{4}(160)]$$

$$S = I$$

$$-20 + \tfrac{1}{4}(160) = 20$$

With the consumption function and investment function as given, we may further illustrate why 160 and only 160 is the equilibrium level of income and output by taking at random any other level of income and output and showing why it is necessarily a disequilibrium level.

Disequilibrium Income and Output

The actual level of output in any period is the result of the decisions of thousands of business persons, and there is no reason to expect their collective decisions to be such as to result precisely in the equilibrium output. In our example, suppose these decisions are such as to result in output of 200. With an output of 200, disposable income will also be 200. The consumption function indicates that income receivers will now spend 170 on consumtion. Adding planned investment of 20 to planned consumption of 170, we have aggregate spending of 190 when output and income are 200. Aggregate spending is clearly insufficient to buy the amount of goods business persons expected to sell. What is required for equilibrium with output of 200 is aggregate spending of 200; what is found is aggregate spending of 190, or a deficiency of 10.

This deficiency of aggregate spending is illustrated in Part A of Figure 4-3 by the difference in height between the aggregate spending function and the guideline. Instead of intersecting the guideline at 200, the aggregate spending function lies below the guideline at 200; the vertical distance between the two lines is the measure of the deficiency of aggregate spending. At this level of output, there is also necessarily disequilibrium between planned saving and planned investment; planned saving is 30 and planned investment is still the unvarying 20. Of the 200 of income earned in the course of producing 200 of output, the 30 that income receivers choose *not* to spend on consumer goods is greater than the 20 that business persons choose to spend on investment goods. In Part A of Figure 4-4, the distance of the saving function above the investment function is the measure of the deficiency of aggregate spending at this level of output.

The equations also indicate that income and output of 200 is a disequilibrium level. Here we do not solve to find what level of Y is the equilibrium level but instead assume a given level of Y and find whether it is the equilibrium level. In the following set of equations, we know that the equilibrium level of Y must be that level of Y which satisfies Equation [2]. Equations [3]

and [4] show that the equilibrium level could not be 200.

$$Y = C + I \quad [1]$$
$$Y = 20 + \tfrac{3}{4}Y + 20 \quad [2]$$
$$200 \neq 20 + \tfrac{3}{4}(200) + 20 \quad [3]$$
$$200 \neq 170 + 20 \quad [4]$$

In terms of saving and investment, the equilibrium level of Y is that level of Y which satisfies Equation [2] in the following set of equations. Could the equilibrium level be 200? Equations [3] and [4] show that it could not.

$$S = I \quad [1]$$
$$-20 + \tfrac{1}{4}Y = 20 \quad [2]$$
$$-20 + \tfrac{1}{4}(200) \neq 20 \quad [3]$$
$$30 \neq 20 \quad [4]$$

With each level of output up to 280 being supplied by business at the same price level of 1, the combination of an aggregate output of 200 and aggregate spending of 190 can mean only one thing: Business as a whole finds its inventories of goods 10 greater than it had planned.[16] If output is maintained at the 200 level period after period and the aggregate spending function remains as given ($20 + \tfrac{3}{4}Y + 20$), business persons will experience an unplanned or involuntary addition of 10 to inventories in each period. Sooner or later, in order to get inventories down to a lower, desired level, business persons will lay off workers and cut back output. This in turn will cause income to fall as fast as output. Once output and income are reduced to 160, equilibrium will be restored; aggregate spending will equal aggregate output, and planned saving will equal planned investment.

To consider another disequilibrium situation, suppose that business persons err in the opposite direction and estimate that they can sell only 120 in output. If output is 120, income will be 120; if income is 120, planned consumption will be 110. Assuming an unvarying 20 of planned investment, aggregate spending will be 110 + 20, or 130, 10 in excess of aggregate output of 120. In Figure 4-3, with output at 120, the aggregate spending function at 130 is 10 above the guideline, and its vertical distance above the guideline is a measure of the excess in aggregate spending, just as its vertical distance below the guideline at output of 200 was a measure of the deficiency in aggregate spending. Planned saving is 10, and planned investment is 20. The 10 of income that income receivers choose *not* to spend for consumer goods is less than the 20 business persons choose to spend for investment goods. The excess of planned investment over planned saving means that aggregate spending must be greater than aggregate output by the amount of this excess. In Figure 4-4 the distance of the investment function above the saving function is the measure of the excess of aggregate spending at this level of output.

The equations also indicate that the income and output level of 120 is a disequilibrium level. The solution to Equation [2] would give us the equilibrium level, and Equations [3] and [4] show that this level could not be 120.

$$Y = C + I \quad [1]$$
$$Y = 20 + \tfrac{3}{4}Y + 20 \quad [2]$$
$$120 \neq 20 + \tfrac{3}{4}(120) + 20 \quad [3]$$
$$120 \neq 110 + 20 \quad [4]$$

Similarly, in terms of saving and investment, the solution to Equation [2] below would give us the equilibrium level. Again Equations [3] and [4] show that this equilibrium level could not be 120.

$$S = I \quad [1]$$
$$-20 + \tfrac{1}{4}Y = 20 \quad [2]$$
$$-20 + \tfrac{1}{4}(120) \neq 20 \quad [3]$$
$$10 \neq 20 \quad [4]$$

In each period during which output remains at 120 and spending at 130, there must be an unplanned decrease of 10 in inventories held

[16] Planned investment of 20 may include a planned increase in inventories. Perhaps plans call for 15 of net investment in plant and equipment and 5 in additional inventories. The result above would thus become an addition of 15 to plant and equipment and 15 to inventories, the planned addition of 5 plus the unplanned addition of 10.

by business persons.[17] Sooner or later, in order to stop this unplanned drain on inventories, business persons will hire more workers and expand output. If they raise output to the 160 level, equilibrium will be restored.

Investment—Planned Versus Realized

We have examined three levels of income and output for our simple economy, of which one (160) was the equilibrium level and the other two (120 and 200) disequilibrium levels. Since income and output are flows, these three levels must all be amounts corresponding to specific time periods. The national income accountant who seeks to measure income and output for these three time periods (the order of which here has no relevance) would summarize the data for the periods as follows:

	Realized	Realized	Realized
	$C + S =$	$Y = C + I$	$S = I$
Period A	$140 + 20 =$	$160 = 140 + 20$	$20 = 20$
Period B	$170 + 30 =$	$200 = 170 + 30$	$30 = 30$
Period C	$110 + 10 =$	$120 = 110 + 10$	$10 = 10$

Notice that the figures in the accountant's identities tell us nothing about the *planned* investment of business persons and nothing about the equilibrium or disequilibrium of the economy at each of these income and output levels. The accountant's identities show only what income and output actually were, how the actual income was divided into realized consumption and realized saving, how the actual output was divided into realized consumption and realized investment, and, from these, the identity between realized saving and realized investment. Whether realized investment is equal to, less than, or greater than planned investment cannot be found from the accountant's identities.

Unlike the accountant, the economist seeks to determine at what level of output the economy will be in equilibrium. We assume that the economist knows what planned consumption spending and planned investment spending will be at each level of income and output and that he or she therefore knows the aggregate spending function ($C + I$) for the economy. In contrast to the accountant's identities, our economist uses a set of equations that show planned consumption and planned investment for the actual income and output levels in each of these time periods. Realized saving and realized investment figures for each period are repeated in Table 4-1 for easy reference.

Comparing the economist's equations with the accountant's identities, we find, as before, that only when income and output are 160 does planned investment of 20 correspond with realized investment. When income and output are 200, business persons discover that, contrary to their plans for investment of 20, realized investment is 30 (consumed output is 170, and the remainder of output or the unconsumed portion of output, 30, equals realized investment). Similarly, when income and output are 120, business persons discover, again contrary to their plans for investment of 20, that realized investment is 10 (consumed output is 110, and unconsumed output of 10 equals realized investment). Since realized investment may be described as the sum of planned and unplanned investment, and since planned investment is the constant 20, the economist may also express this in equation form as follows:

	Planned Investment	+	Unplanned Investment	=	Realized Investment	=	Realized Saving
Period A	20	+	0	=	20	=	20
Period B	20	+	10	=	30	=	30
Period C	20	+	−10	=	10	=	10

[17] If plans had called for net investment of 10 in plant and equipment and 10 in inventories per time period, the addition to inventories would be 0. If plans had called for net investment of 20 in plant and equipment and no addition to inventories per time period, the results would be net investment of 20 in plant and equipment and −10 (a decrease) in inventories. In both cases, realized investment, or unconsumed output would be 10 (120 − 110).

TABLE 4-1

	PLANNED $C + S = Y = C + I$ PLANNED	PLANNED $S = I$	REALIZED $S = I$
Period A	$140 + 20 = 160 = 140 + 20$	$20 = 20$	$20 = 20$
Period B	$170 + 30 = 200 > 170 + 20$	$30 > 20$	$30 = 30$
Period C	$110 + 10 = 120 < 110 + 20$	$10 < 20$	$10 = 10$

In graphic form, the amounts of unplanned investment may be identified in Part A of Figures 4-3 and 4-4 as the difference between realized investment and planned investment at each of the three levels of income and output.[18] When unplanned investment is 10, the excess of realized investment over planned investment amounts to an unplanned increase in inventories of 10; some of the goods produced simply are not sold and remain in inventory, even though the producers do not want them to. When unplanned investment is −10, on the other hand, the excess of planned investment over realized investment amounts to an unplanned decrease in inventories of 10.

In short, if there is any unplanned investment, planned investment will not equal realized investment, and the economy will therefore be at a disequilibrium level of output. However, the identity between realized saving and realized investment is just as consistent with positive or negative unplanned investment as it is with zero unplanned investment.[19] The accounting identities, therefore, can tell us nothing about whether there is equilibrium or disequilibrium in a particular time period and therefore nothing about whether the level of income and output will rise or fall in succeeding time periods. To determine this we need equations that show the relation between planned saving and planned investment.

Hence we must work with definitions of saving and investment that at first appear to be contradictory. In the one definition, saving and investment (realized) are necessarily equal in any time period; in the other definition, saving and investment (planned) are not necessarily equal and in fact are typically unequal in any time period. In the introduction to national income accounting in Chapter 2, we used only the first definition of saving and investment and did not have to distinguish between planned and realized. From now on, however, they must be so distinguished. We must avoid saying in the same breath that saving and investment can only be equal and that saving and investment can be unequal. We can, however, flatly and unambiguously say that realized saving and realized investment can only be equal but that planned saving and planned investment can be unequal. Throughout the remainder of this book, all references to consumption, saving, and investment not specifically designated as realized quantities should be understood to be planned quantities.

[18] Realized investment at any level of output is the difference between the C curve and the guideline in Part A of Figure 4-3 and the difference between the horizontal axis and the S curve in Part A of Figure 4-4. Planned investment is the difference between the C curve and the $C + I$ curve in Part A of Figure 4-3 and the difference between the horizontal axis and the I curve in Part A of Figure 4-4.

[19] A discrepancy can appear only between planned investment and realized investment and not between planned saving and realized saving during a particular time period, because it has been implicitly assumed that income receivers succeed in saving the amount of income they plan to save at each level of income. One can drop this assumption and consider disequilibrium situations in which business persons always succeed in investing the amount they plan to invest while income receivers fail to save the amount they plan to save at any given level of income. Disequilibrium will then appear as a result of planned saving being greater than or less than realized saving.

APPENDIX
A NOTE ON THE PERFECTLY INELASTIC AGGREGATE DEMAND FUNCTION

At the beginning of this chapter, we indicated in general terms the rationale for a perfectly inelastic *AD* curve. In Figures 4-3 and 4-4, we related the *AD* curve in Part B to the curves in Part A from which the *AD* curve was derived. However, the nature of the relationship is not very clear when the *AD* curve is perfectly inelastic as in these figures. Although all of our work with the Keynesian model until Chapter 17 will employ a perfectly inelastic *AD* curve, the nature of the relationship between the *AD* curve in Part B and its determinants in Part A of Figures 4-3 and 4-4 may be seen more clearly in the case of a downward-sloping *AD* curve. The purpose of this appendix is to clarify the relationship in question by showing what lies behind a downward-sloping *AD* curve within the framework of the simple model of Figures 4-3 and 4-4.

The conditions that must be satisfied to obtain a downward-sloping *AD* curve within this model are that in Part A of Figure 4-3, the position of the $C + I$ curve must vary inversely with the price level because either C or I or both respond to changes in the price level in this way. The same conditions in Part A of Figure 4-4 are that the position of the S curve must vary directly with the price level, or the position of the I curve must vary inversely with the price level, or both. What this says, directly in Figure 4-3 and indirectly in Figure 4-4, is that, other things being equal, the total amount of consumer spending and investment spending at any level of real income, Y, must be lower at a higher price level and higher at a lower price level.[1]

[1] The way that aggregate spending varies with the price level must be carefully distinguished from the way that aggregate spending varies with different rates of inflation and, beyond this, with different expected rates of inflation. The comparison here is between different absolute price levels and not between different rates of price level change. The actual rate of inflation and the expected rate of inflation bring quite different considerations into the analysis.

To illustrate the required conditions, Part A of Figure 4-5 shows three consumption functions labeled $C_{0.5}$, $C_{1.0}$, and $C_{1.5}$ in which the subscripts refer to the three price levels, 0.5, 1.0, and 1.5, shown specifically in Part B. (To avoid complicating the figure unnecessarily, it is assumed that I does not vary with P; as before, it is a constant 20 at all levels of Y.) As $1.5 > 1.0 > 0.5$, the $C_{1.5}$ curve lies below the $C_{1.0}$ curve and the $C_{1.0}$ curve lies below the $C_{0.5}$ curve. The vertical distance between the C curves is 15 or a change in P of 0.5 shifts the consumption function vertically by 15.

Accompanying the C curves are the $C + I$ curves, each 20 above its respective C curve. With $P = 1.5$, $C_{1.5} + I$ intersects the 45° line at $Y = 100$ or aggregate spending will be just that needed to remove output of 100 from the market, if output happens to be 100. By the same reasoning, if $P = 1.0$, aggregate spending will be just that needed to remove output of 160 from the market, if output happens to be 160. For the third case of $P = .05$, the same may be said for the output level of 220. These combinations of P and Y identify the three points, J, K, and L, in Part B of the figure. Connecting these points yields the downward-sloping *AD* curve shown. Each point along this curve indicates the P at which *AD* will be just the amount needed to purchase the Y found by dropping a perpendicular from that point to the Y axis. So far as the *AD* curve is concerned, each Y level identified in this way is an amount of Y that can be sold at the specified P. This much we know from Part A of the figure which gives all the information needed to draw the *AD* curve in Part B.

What Part A of the figure and the *AD* curve derived therefrom cannot tell us is what the equilibrium combination of P and Y will be. This cannot be discovered just by knowing how much can be sold at each price level. One must also know how much will be supplied at each

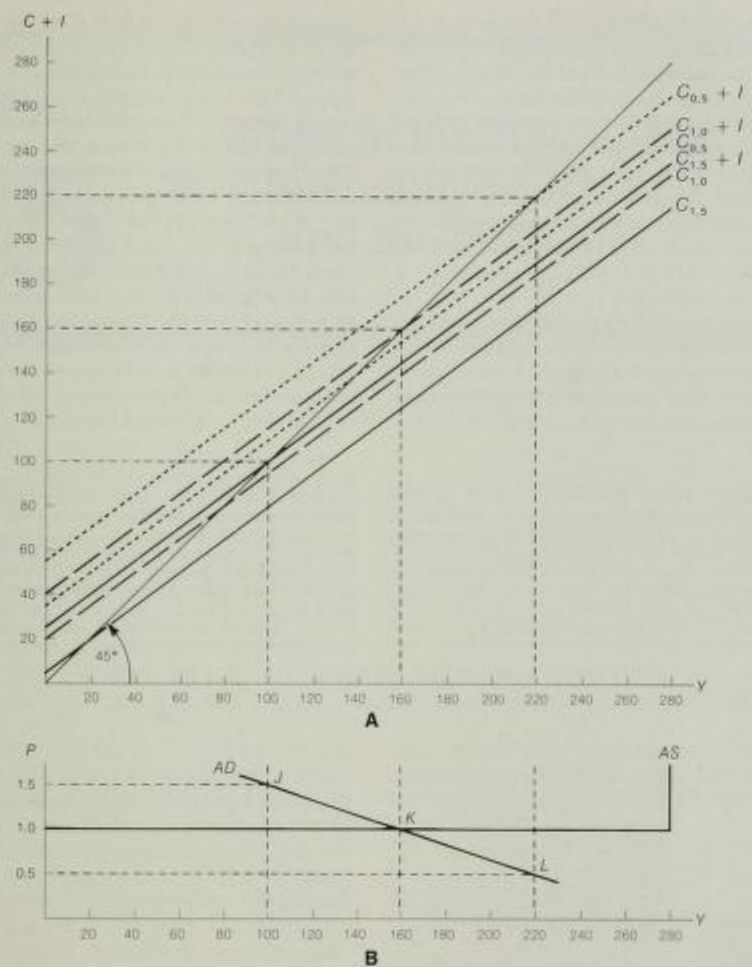


FIGURE 4-5
Derivation of a downward-sloping aggregate demand curve

price level, and that information is provided by the aggregate supply curve. Once the *AS* curve is inserted, the answer to what the equilibrium *P* and *Y* will be is given by the intersection of the *AD* and *AS* curves: in the present illustration, *P* of 1.0 and *Y* of 160. With the downward-sloping *AD* curve as shown, if *Y* were 100, *AD* would remove the 100 from the market at a price level of 1.5. However, this is not an equilibrium. As shown by the *AS* curve, firms are

willing to sell more than 100 at a P less than 1.5 and under competitive conditions P will be bid down. The decrease in P will continue to 1.0 at which level there is equilibrium as shown by the fact that $AD = AS$. Similarly, if Y were 220, AD would remove the 220 from the market at a price level of 0.5. Here again there is not an equilibrium. Firms will not supply 220 (or in this case any other amount) at a P below 1.0. In this case buyers bid up P. The increase in P will continue to 1.0 at which price level there is equilibrium: $AD = AS$.

We have worked through this exercise to clarify the relationship between the AD curve in Part B and its determinants in Part A.[2] The relationship comes out more clearly in the present case of a downward-sloping AD curve than in the case of the perfectly inelastic AD curve. To obtain a downward-sloping AD curve in the present model requires that the position of the $C + I$ curve in Part A vary inversely with P. As we will see in Chapter 17, an argument may be advanced to support this inverse relationship, but this argument is not part of the simple Keynesian model with which we will be dealing in this part of the book.[3] In this simple model a perfectly inelastic AD curve is appropriate (and offers the advantage of relatively uncomplicated figures). A curve of this kind will be assumed in all our work with the Keynesian model until Chapter 17, where the wealth effect and other effects that produce a downward-sloping AD curve are brought into the analysis in the course of producing an extended Keynesian model.

[2]This exercise has been based on the $C + I$ approach of Figure 4-3, but the same downward-sloping AD curve may be derived through the S, I approach of Figure 4-4. In Part A shift the S curve downward for each decrease in P by the same amount that the C curve in the figure here was shifted upward for each decrease in P. S will equal I at Y of 100 with P of 1.5, at 160 with P of 1.0, and at 220 with P of 0.5.

[3]The argument in question, in a word, is that a change in the price level has an effect on the public's *wealth*. The public holds certain assets such as money and government bonds which are fixed in dollar terms. A fall in the price level increases the real value of a given face amount of these assets, makes the owners feel wealthier, and leads them to devote a larger fraction of their current real income to the purchase of goods and services. Thus, other things being equal, the lower is P, the higher the position of the C curve. From this straightforwardly follows a downward-sloping AD curve as in the figure above.

chapter 5 12

Shifts in the Aggregate Spending Function and the Multiplier

In the preceding chapter we saw that output may fluctuate above and below the equilibrium level as business persons err in overproducing or underproducing. We also saw that they react to these errors in such a way as to make output move toward, if not actually to attain, that single equilibrium level of output which in the simple model is determined solely by the position of the aggregate demand function.

SHIFTS IN THE AGGREGATE SPENDING FUNCTION

From one time period to the next, it is not only possible but probable that the perfectly inelastic aggregate demand function in Part B of Figures 4-3 and 4-4 will shift to the right or the left as shifts occur in the aggregate spending function of Part A of Figure 4-3 or, what is the same thing, as shifts occur in the saving or investment functions of Part A of Figure 4-4. In our two-sector economy the aggregate spending function is the sum of the consumption function and the investment function. While it is possible for either or both of these functions to shift from one time period to the next, most observers agree that the consumption function is relatively stable and the investment function relatively unstable. The relative stability of the consumption function does not necessarily mean that the actual amount of consumption expenditures is relatively stable, since this amount will change with every change in the level of income. It does mean, however, that the amount of consumption expenditures *at any given level of income* is relatively stable and that an initial change in the level of income itself is typically the result of a shift in the investment function.[1] In terms of

[1]Given a stable consumption function of $C = 20 + \frac{3}{4}Y$, C can change only as a result of a change in Y. If Y were 100, C would be 95; if Y were 120, C would be 110. *Consumption expenditures* are not stable, for they change with the level of income; but, under these circumstances, the *consumption function* is perfectly stable. An unstable consumption function would show change in C with no change in Y, which could result only from a change in the value of the constant 20 or of the constant 3/4. Thus, if Y is 120 and stays at 120, C can still change from one period to the next if the constant 20 rises to, say, 25.

Part A of Figure 4-3, this means simply that the entire C curve does not bounce up and down over the short run. The $C + I$ curve does fluctuate over the short run, but this is due primarily to instability in the investment component rather than in the consumption component. For this reason, the analysis in this chapter will concentrate on changes in the equilibrium level of income and output that result from shifts in the investment function. It should be noted, however, that a parallel analysis applies to shifts in the consumption function.

To begin, assume that, due to an improvement in business expectations, investment expenditures rise permanently from 20 to 30 per time period, an amount that is, as before, the same at all levels of output. Part A of Figure 5-1 illustrates this shift in the investment function. The curves labeled C and $C + I$ are the same as those found in Part A of Figure 4-3. To indicate the increase from 20 to 30 in investment expenditures, we simply add vertically to the curve labeled $C + I$ the 10 of additional investment and label this new aggregate spending function $C + I + \Delta I$. The curves labeled S and I in Part B of Figure 5-1 are the same as those found in Part A of Figure 4-4; we simply add vertically to the curve labeled I the 10 of additional investment and label this new investment function $I + \Delta I$. Lastly, the AS curve in Part C is the same as that found in Part B of both Figures 4-3 and 4-4.

With the aggregate spending function, $C + I$, the equilibrium level of output was 160. With the new aggregate spending function 10 higher than the old, we might expect the new equilibrium level of output to be also 10 higher, or an increase from 160 to 170. But 170 is not the new equilibrium level for at 170 aggregate spending exceeds aggregate output and investment exceeds saving. The equilibrium level, therefore, must be greater than 170. Part A of Figure 5-1 reveals that, with the higher aggregate spending function, aggregate spending and aggregate output are equal only at an output of 200, and Part B reveals that, with the higher investment function, saving and investment are equal only at an output of 200. Thus, given that the increase of 10 in investment increases the level of aggregate demand from 160 to 200, there is a shift in the aggregate demand curve of Part C from its original position at Y of 160 shown by the curve labeled AD_I to a new position at Y of 200 shown by the curve labeled $AD_{I+\Delta I}$. The fact that the increase of 10 in investment has raised the equilibrium level of income and output not merely by 10 but by 40, perhaps surprising at first glance, will be explained in the following pages.

A Temporary Shift in the Aggregate Spending Function

The new equilibrium level in Figure 5-1 was established as the aggregate spending function shifted from $C + I$ to the higher level of $C + I + \Delta I$. We have assumed that this is a permanent shift. However, to understand better the implications of a *permanent* shift, let us first examine the implications of a *temporary* shift in the function. Unlike the results shown in Figure 5-1, if the rise in investment is a temporary one, the rise in income and output will also be temporary. When investment subsequently drops back to its original level, income and output will also eventually drop back to their original level. The aggregate demand curve in Part C of Figure 5-1 is positioned at Y of 160 before and after the temporary increase in investment.

Changes in the level of income and output take place over time. To trace the process set into motion by a temporary increase in investment spending, we may split up the time interval required for the system to reestablish equilibrium into a series of shorter, numbered time periods. Let us assume that in period 1 we have the original equilibrium position described in the previous chapter and shown in Figure 5-1; aggregate output is 160 and aggregate spending, the sum of consumption spending of 140 and investment spending of 20, is also 160. Then in period 2 we upset this equilibrium by introducing an increase in investment spending

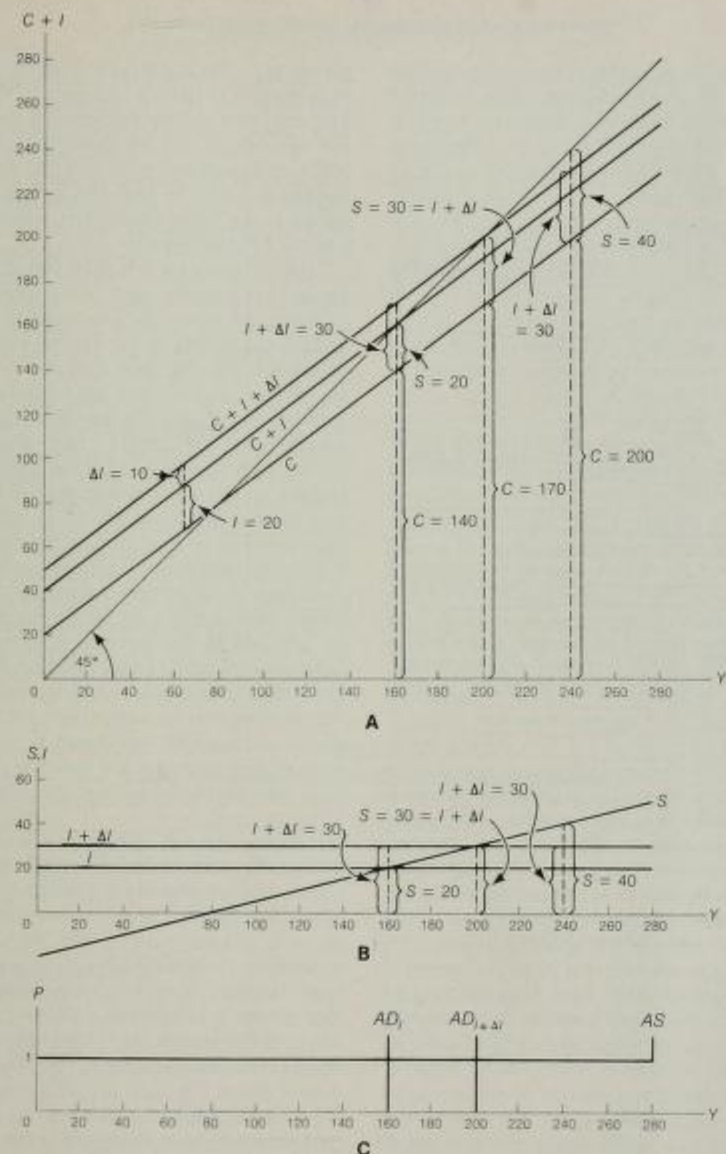


FIGURE 5-1
The effect of a change in investment on the equilibrium level of income

X — read

of 10, or a rise in investment spending from 20 to 30. We assume that business persons do not attempt to anticipate changes in the demand for their output but instead follow a simple rule of thumb of producing in each period an output equal to their sales in the preceding period. Given this behavior by business persons, output in period 2 will be equal to sales of period 1; that is, output will be 160. Aggregate spending in period 2, however, is found to be 170, for in this period we have the increase of 10 in investment spending. Inventories serve as a buffer—the *excess* of aggregate spending of 10 is absorbed in period 2 by an unplanned decrease of inventories.

In period 3, business persons expand aggregate output to 170, the figure for total sales in period 2. Output of 170 in period 3 generates disposable income of 170 during this period. Given the consumption function, $C = 20 + 3/4Y$, consumption spending in period 3 will be 147.5 or 7.5 greater than in period 2. If now in period 3 investment demand drops back to its original level of 20 following its temporary rise to 30 in period 2, we find in period 3 consumption spending of 147.5 and investment spending of 20, or aggregate demand of 167.5. Since output in period 3 is 170, there is in this period a *deficiency* of aggregate spending of 2.5, which is reflected in an unplanned increase in inventories. In period 4, business persons reduce aggregate output to 167.5, the total for sales in period 3. This means a corresponding decline in income to 167.5, and the consumption function indicates that consumption in turn will be 145.6 in period 4. Spending in period 4 is accordingly this 145.6 plus 20 for investment or an aggregate of 165.6. Since output in period 4 is 167.5, there is again a deficiency of spending now equal to 1.9, which again is reflected in an unplanned increase in inventories. In period 5, there is a further reduction in output and again a deficiency of spending but one smaller than that of period 4. In this way, the level of output declines period by period until period n, assumed to be the last in what is actually an infinite number of periods. In period n, output is 160, income is correspondingly 160, consumption spending is 140, and investment spending is 20—or aggregate spending of 160 is equal to aggregate output of 160. The system has returned to the same equilibrium position from which we started in period 1.

Table 5-1 gives the period-by-period detail of the process just described. In this table, C and I indicate the values for consumption and investment spending in the original equilibrium of period 1. The change in consumption spending between period 1 and any following period is shown by ΔC, and the change in investment spending between period 1 and any following period is shown by ΔI. Total consumption spending in any period is then given by $C + \Delta C$, total investment spending by $I + \Delta I$, and aggregate spending by their sum. In the same way, Y indicates aggregate output in period 1, ΔY the change in output between period 1 and any other period, and $Y + \Delta Y$ the aggregate output for any period. The last two columns in the table show for each period *realized* investment, which equals saving, and *planned* investment. Planned investment for each period is the same as $I + \Delta I$. Realized investment or saving (or unconsumed output) in any period is the difference between that period's output and its consumption, or $(Y + \Delta Y) - (C + \Delta C)$.

In any time period, aggregate spending may be equal to, greater than, or less than aggregate output. To say the same thing in different words, in any time period, planned investment may be equal to, greater than, or less than realized investment. The level of output is an equilibrium level in period 1, for in this period we find aggregate spending of 160 just equal to aggregate output of 160 and planned investment of 20 just equal to realized investment of 20. The equilibrium of period 1 is upset in period 2 by the rise in planned investment, which is a rise in aggregate spending; and equilibrium is not restored until period n, for in all intervening periods we

find aggregate spending either greater than or less than aggregate output or, what is the same thing, planned investment either greater than or less than realized investment.

We find disequilibrium through all these periods, despite the fact that the cause of disequilibrium, the rise in investment spending, is limited to period 2 alone. This one-period rise in investment spending, however, produces the changing level of output in all these later periods by initiating a series of changes in consumption spending starting in period 3. Thus, ΔC of 7.5 in period 3 results from ΔY of 10 in period 3, which results from ΔI of 10 in period 2. Similarly, ΔC of 5.6 in period 4 results from ΔY of 7.5 in period 4, which results from ΔC of 7.5 in period 3, which results from ΔY of 10 in period 3, which results from ΔI of 10 in period 2. In other words, the ΔI of period 2 initiates a process in which ΔC in each period is 3/4 (the MPC) of ΔY in that period. Furthermore, given that ΔC is equal to 3/4 of ΔY of that period and that ΔY of each period is equal to $\Delta C + \Delta I$ of the preceding period, the fact that the increase in investment, ΔI, is limited to period 2 alone means that ΔY and ΔC become smaller each period until eventually, in period n, ΔC becomes zero.[2] Since ΔI is also zero in period n, ΔY becomes zero, and the level of income and output is back to the equilibrium found in period 1, or 160.[3]

Although Table 5-1 shows that the equilibrium level of output established in period n is the same as that of period 1, it is important to note

[2]For a one-period increase in investment spending of $1, the *differences* between each period's aggregate spending and spending in the original period are given by the series

$$1, c, c^2, c^3, c^4, \ldots, c^n$$

Since c is less than 1, the differences become smaller and smaller. After the passage of an infinite number of time periods represented by n, c^n becomes infinitely small, so that income returns to its original equilibrium level.

[3]Actually, the cumulative addition to investment spending does not occur in period 2 as is suggested by the column headed ΔI. An increase in investment of 10 was planned for period 2, but the realized increase in investment was zero. As shown by the next to last column, realized investment in period 2 is 20, the same as in period 1. The *realized* increases in investment may be identified in this column as the differences between the indicated values and 20; that is, 2.5 in period 3, 1.9 in period 4, and so forth. The sum of these changes will total 10, equal to the increase in planned investment of 10 shown in period 2.

TABLE 5-1
A temporary (one-period) increase in investment spending

PERIOD	C +	I +	ΔC +	ΔI =	AGGREGATE SPENDING	$\gtreqless$	AGGREGATE OUTPUT	= Y +	ΔY	REALIZED INVESTMENT	PLANNED INVESTMENT
1	140 +	20 +	0.0 +	0 =	160.0	=	160.0	= 160 +	0.0	20.0	20
2	140 +	20 +	0.0 +	10 =	170.0	>	160.0	= 160 +	0.0	20.0	30
3	140 +	20 +	7.5 +	0 =	167.5	<	170.0	= 160 +	10.0	22.5	20
4	140 +	20 +	5.6 +	0 =	165.6	<	167.5	= 160 +	7.5	21.9	20
5	140 +	20 +	4.2 +	0 =	164.2	<	165.6	= 160 +	5.6	21.4	20
6	140 +	20 +	3.2 +	0 =	163.2	<	164.2	= 160 +	4.2	21.0	20
7	140 +	20 +	2.4 +	0 =	162.4	<	163.2	= 160 +	3.2	20.8	20
8	140 +	20 +	1.8 +	0 =	161.8	<	162.4	= 160 +	2.4	20.6	20
9	140 +	20 +	1.3 +	0 =	161.3	<	161.8	= 160 +	1.8	20.5	20
⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮
n	140 +	20 +	0.0 +	0 =	160.0	=	160.0	= 160 +	0.0	20.0	20
$n + 1$	140 +	20 +	0.0 +	0 =	160.0	=	160.0	= 160 +	0.0	20.0	20
			30.0	10					40.0		

that the *cumulative* addition to income and output over the time interval in which the process works itself out is, in the present case, four times the size of the initiating increase in investment spending in period 2. This cumulative addition to income and output is 40, as shown at the bottom of the column headed ΔY in Table 5-1. The 40 is composed of cumulative additions of 30 to consumption and 10 to investment, as shown at the bottom of the columns headed ΔC and ΔI in the table.[3] Thus, although the one-period injection of extra investment spending does not lift income and output to a permanently higher level, the *cumulative* effect in the present case is a flow of extra income and output four times the amount of that one-period injection of investment spending.

A Permanent Shift in the Aggregate Spending Function

If the aggregate spending function shifts upward and remains at the new higher level period after period, the original equilibrium level of income and output will be replaced by a new, higher equilibrium level. This is the result shown in Figure 5-1. The original position of the aggregate demand curve is at Y of 160 as shown in Part C by the *AD* curve erected at that level of Y. The intersection between the *AD* and the *AS* curves indicates an equilibrium level of income and output of 160 and an equilibrium price level of 1. After the permanent upward shift of 10 in the aggregate spending function, the new position of the aggregate demand curve is at Y of 200 as shown by the *AD* curve erected at that level of Y. The intersection between the new *AD* curve and the *AS* curve indicates an equilibrium level of income and output of 200 and an equilibrium price level of 1.

To describe the process by which the system moves to a higher equilibrium level of output as a result of such a sustained increase in aggregate spending requires only that we extend the description of the process for a temporary increase in aggregate spending. In Table 5-1, ΔI of 10 in period 2 called forth ΔY of 10 in period 3, which in turn resulted in ΔC of 7.5 in period 3. Because we then assumed that investment spending returned to its original level in period 3, ΔY of period 4, equal only to ΔC of 7.5 of period 3, dropped below ΔY of period 3, which was equal to ΔI of 10 of period 2. In the present case, however, with a permanent rise in investment spending of 10, we have in period 3 ΔI of 10 as well as ΔC of 7.5, so that ΔY of period 4, equal now to ΔC of 7.5 plus ΔI of 10, is 17.5, greater than ΔY of 10 in period 3. From ΔY of 17.5 in period 4, we get ΔC of 13.1 in period 4, so that ΔY in period 5 is 23.1, the sum of ΔC of 13.1 plus the constant ΔI of 10. Thus, with investment sustained at the new higher level, ΔY in each period continues to rise above ΔY of the preceding period until a new equilibrium is established with ΔY of 40.

Table 5-2 records the period-by-period detail of this process. The column headings are the same as those in Table 5-1. ΔI starts out at zero in period 1, becomes 10 in period 2, and remains 10 in each succeeding period. As before, ΔY for any period is the sum of ΔI plus ΔC for the preceding period. In comparing the value of ΔY period by period, note that as ΔY becomes larger and larger, the change in ΔY becomes progressively smaller and in period n becomes zero. In this period ΔY stabilizes at 40, and the sum of $Y + \Delta Y$ stabilizes at 200. With aggregate output at 200, income is 200, and with income at 200, the consumption function, $C = 20 + 3/4Y$, indicates consumption spending of 170. Investment spending is given at 30, so aggregate spending in period n is 200 and equal to aggregate output. Similarly, as shown in the last two columns of Table 5-2, in period n realized investment reaches 30 and is equal to planned investment of 30. Viewed in either way, in period n a new equilibrium is established. With this equilibrium at Y of 200, the aggregate demand curve is correspondingly positioned at Y of 200.

THE MULTIPLIER—A SHIFT IN THE AGGREGATE SPENDING FUNCTION

In terms of Figure 5-1 and Table 5-2, a permanent upward shift in the aggregate spending function results in a movement of income and output to a new equilibrium level that is higher than the original equilibrium level by some multiple of the upward shift in the aggregate spending function. The value of this multiple is known as the *multiplier* and represents the number by which the shift in the aggregate spending function must be multiplied to determine the change in the level of income and output required to establish a new equilibrium.[*] Under present assumptions, this multiple is 4.

[*]Though made famous by the role it plays in Keynes's *General Theory*, the term itself was coined by another British economist, R. F. Kahn. Kahn's multiplier was an *employment* multiplier, measuring the ratio of the increment of total employment associated with a given increment of employment in the capital goods industries. Keynes's multiplier is an *investment* multiplier, the ratio of the increment to total income associated with a given increment in investment. (See John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936, pp. 113–15.)

Why do Figure 5-1 and Table 5-2 show a multiplier of 4 and not some other? The reason is that income receivers choose to spend on consumption 3/4 of any change in income (MPC = 3/4) or that they choose to save, or not spend on consumption, 1/4 of any change in income (MPS = 1/4). With MPS = 1/4, it is only when income and output have risen by 40 that income receivers will devote an additional 10 of their higher income to saving. Then $\Delta S = \Delta I$, $10 = 10$, and a new equilibrium level of income and output is established.

Thus, with any given shift in the aggregate spending function, the change in income required to reestablish equilibrium is entirely dependent on the value of the MPC or the MPS. For example, still assuming that $\Delta I = 10$, we can simply determine the new equilibrium if MPC = 4/5 and MPS = 1/5. Instead of ΔY of 40 as before, we will now have ΔY of 50, for it is only when ΔY is 50 that ΔS will be 10 and therefore equal to ΔI of 10. Accordingly, until Y

TABLE 5-2
A permanent increase in investment spending

PERIOD	$C + I + \Delta C + \Delta I =$	AGGREGATE SPENDING	$\gtreqless$	AGGREGATE OUTPUT	$= Y + \Delta Y$	REALIZED INVESTMENT	PLANNED INVESTMENT
1	140 + 20 + 0.0 + 0 =	160.0	=	160.0	= 160 + 0.0	20.0	20
2	140 + 20 + 0.0 + 10 =	170.0	>	160.0	= 160 + 0.0	20.0	30
3	140 + 20 + 7.5 + 10 =	177.5	>	170.0	= 160 + 10.0	22.5	30
4	140 + 20 + 13.1 + 10 =	183.1	>	177.5	= 160 + 17.5	24.4	30
5	140 + 20 + 17.3 + 10 =	187.3	>	183.1	= 160 + 23.1	25.8	30
6	140 + 20 + 20.5 + 10 =	190.5	>	187.3	= 160 + 27.3	26.8	30
7	140 + 20 + 22.9 + 10 =	192.9	>	190.5	= 160 + 30.5	27.6	30
8	140 + 20 + 24.7 + 10 =	194.7	>	192.9	= 160 + 32.9	28.2	30
9	140 + 20 + 26.0 + 10 =	196.0	>	194.7	= 160 + 34.7	28.7	30
⋮	⋮	⋮		⋮	⋮	⋮	⋮
n	140 + 20 + 30.0 + 10 =	200.0	=	200.0	= 160 + 40.0	30.0	30
$n + 1$	140 + 20 + 30.0 + 10 =	200.0	=	200.0	= 160 + 40.0	30.0	30

has risen by 50 (until ΔY equals 50), aggregate spending will exceed aggregate output and investment will exceed saving, forcing a further rise in income and output. By the same reasoning, if the MPC were 2/3 and the MPS 1/3, the rise in income would be 30.[5]

These and various other combinations of MPC and MPS are all possible and plausible. Although many combinations are not plausible, there is a particular pair of these combinations that can help clarify the multiplier mechanism. One of these is MPC = 0 and MPS = 1. The rise in Y necessary to reestablish equilibrium in this case will be exactly equal to the permanent rise in I. If ΔI in period 2 and in each subsequent period is 10, ΔY in period 3 and in each subsequent period will be 10 also; for, if MPS = 1, ΔS will be 10 as soon as ΔY is 10. Therefore, $\Delta S = \Delta I$ in period 3, and the new equilibrium level is immediately established at a level of income and output exactly 10 above the original level. The increase in investment does not lead to an increase in income larger than itself; for, with MPS = 1, the rise in income of 10 in period 3 does not lead to a rise in consumption spending in period 3. Instead, the income receivers choose to devote all the increase in income of period 3 to saving (MPS = 1). There being no induced rise in consumption spending in period 3, ΔY of period 4 and each subsequent period is simply equal to ΔI of 10 for each such period. Unlike the situation in Table 5-2, in which only a part of the enlarged income stream of period 3 was diverted into saving, here all of ΔY of period 3 leaks out of the spending stream in period 3, and the expansion process ends as quickly as it began. In a formal sense, the value of the multiplier is 1, but this is a far cry from the earlier results in which the multiplier was 4 (MPC = 3/4 and MPS = 1/4).

The other combination is MPC = 1 and MPS = 0. Starting off as before with ΔI of 10 in period 2, and in each subsequent period, none of the additional income of 10 flowing to income receivers in period 3 is diverted from the spending stream. Thus, the entire 10 of ΔY of period 3 appears on the market as spending for consumption goods. In period 3, $\Delta C = 10$, $\Delta I = 10$, and therefore in period 4 $\Delta Y = 20$. In period 4, all ΔY of that period appears as spending for consumption goods so that $\Delta C = 20$, $\Delta I = 10$, and therefore in period 5, $\Delta Y = 30$. No new equilibrium would be possible in this case, and income would rise without limit. Equilibrium requires that $\Delta S = \Delta I$. But since all ΔY of any period is devoted to ΔC and none is diverted to ΔS, ΔS remains zero period after period and can never equal ΔI as required for equilibrium. Period after period, investment exceeds saving, and aggregate spending exceeds aggregate output. In this special situation, we quickly have to drop our assumption of a stable price level because the aggregate demand curve in Part C of Figure 5-1 will before many periods move rightward to the point at which the aggregate supply curve changes from perfect elasticity to perfect inelasticity. At this point a rightward shift in the aggregate demand curve becomes purely inflationary, as it calls forth no further increase in output but only raises the price level. Such is the consequence of the assumption of an increase in I with MPC = 1 and MPS = 0.

In all the examples above, we have assumed that the aggregate spending function shifts upward. Shifts in the opposite direction are equally possible. In such cases, the multiplier works against us instead of for us; it works to produce a multiple contraction of income and output instead of a multiple expansion. Thus, with a downward shift in the aggregate spending func-

[5] While we may ask such questions and give such answers to illustrate the principle, it should be noted that if we were to assume some value other than 3/4 for the MPC (or 1/4 for the MPS) the original equilibrium would not have been 160. With an MPC of 4/5, the original equilibrium income and output would have been 200, equal to the aggregate spending function, $20 + \frac{4}{5}Y + I$, in which I equals 20. From this original equilibrium, ΔI of 10 would result in ΔY of 50 before equilibrium was restored with ΔS of 10 equal to ΔI of 10. ΔY would be 50 as described above, but the rise in Y would have been from an original equilibrium with $Y = 200$ to a new equilibrium with $Y + \Delta Y = 250$.

tion, the question to be answered is not how much income will rise before equilibrium is restored but how much it will fall before equilibrium is restored. With MPC = 3/4 and MPS = 1/4, if investment spending falls from 20 to 10 ($\Delta I = -10$), the *drop* in income and output necessary to restore equilibrium will be 40. The reasoning is the same as before. Income will drop until saving again equals investment. Since investment is reduced from 20 to 10, or by −10, saving must be reduced from 20 to 10, or by −10, to restore equilibrium. Given MPS = 1/4, only when income is reduced by 40 will saving be reduced by 10. As income receivers increase saving by 1/4 of any addition to income, so they decrease saving by 1/4 of any reduction in income. From the original equilibrium of 160, the system reaches a new equilibrium level with income and output reduced by 40 to 120. The aggregate demand curve shifts leftward from its original position at Y of 160 to a new position at Y of 120. The downward multiplier is 4 for the same reason that the upward multiplier is 4: because MPC = 3/4 and MPS = 1/4. If we assume different values for the MPC and the MPS, the results for a downward shift in the aggregate spending function parallel those just described, the only difference being the size of the multiplier. With MPC = 2/3 and MPS = 1/3, a downward shift in the aggregate spending function of 10 ($\Delta I = -10$) will produce a decline in income and output of 30.

Since the MPC or the MPS determines the multiplier and since the multiplier determines the size of the increase or decrease in income and output that will follow any given upward or downward shift in the aggregate spending function, the practical importance of the MPC and the MPS is great. Given the variability of that portion of the aggregate spending function made up of investment spending, the degree of instability of the entire economic system depends to some extent on the values of the MPC and the MPS. As business spending for plant and equipment goes up in one period and down in another, there is a direct impact on the level of income and output within each period. This in itself is a source of instability. Yet whatever the variability of investment spending, a relatively low MPC and a relatively high MPS will tend to produce less instability in the economy than will a relatively high MPC and a relatively low MPS. Of the two extreme cases we discussed earlier, that in which MPC = 1 and MPS = 0 will produce extreme instability in income and output, since any variability in investment spending from one period to the next will be greatly magnified by continuously rising induced consumption spending. At the other extreme, when MPC = 0 and MPS = 1, the instability in income and output will be far less, because the variability of investment spending from one period to the next will not be magnified at all by induced consumption spending.

Although any explanation of fluctuations in the level of income and output involves far more than just the variability of investment spending and the values of the MPC and the MPS, these values still play a vital role in explaining the amplitude of the upward and downward movements in income and output during business cycles. More will be said about this topic in Chapter 19, where the multiplier will be made a specific part of a multiplier—accelerator model of the business cycle.

THE MULTIPLIER—EQUATIONS

To determine the equilibrium level of income and output in the two-sector economy, we used the equation

$$Y = C + I \qquad [1]$$

Once given the consumption function and the investment function, the equation could be

readily solved.

$$Y = 20 + \tfrac{3}{4}Y + 20$$
$$160 = 140 + 20$$

If we retain this same consumption function but assume now an upward shift in the investment function, there will be an increase in the equilibrium level of income and output. Since any increase in Y, ΔY, must be equal to $\Delta C + \Delta I$, we have the following equation, the solution to which gives us the new equilibrium level of income and output. (This equation is the one from which the column headings in Tables 5-1 and 5-2 came.)

$$\begin{aligned} Y = \Delta Y &= C + I + \Delta C + \Delta I \\ 160 + \Delta Y &= 140 + 20 + \Delta C + \Delta I \end{aligned} \qquad [2]$$

Subtracting Equation [1], $Y = C + I$, from Equation [2], we have another equation, the solution to which indicates the *change* in the level of income necessary to produce the new equilibrium level of income.

$$\Delta Y = \Delta C + \Delta I \qquad [3]$$

The consumption function, $C = C_a + cY$, indicates that consumption spending, C, rises or falls by an amount equal to the MPC or c (here 3/4) times the change in income. That is, it says that $\Delta C = c\,\Delta Y$ or, in the present example, $\Delta C = 3/4\ \Delta Y$. Substituting in Equation [3], we have the following:

$$\Delta Y = c\,\Delta Y + \Delta I$$
$$\Delta Y - c\,\Delta Y = \Delta I$$
$$\Delta Y(1 - c) = \Delta I$$
$$\Delta Y = \frac{1}{1 - c}\Delta I$$

and

$$\frac{\Delta Y}{\Delta I} = \frac{1}{1 - c}$$

or

$$\Delta Y = \tfrac{3}{4}\Delta Y + \Delta I$$
$$\Delta Y - \tfrac{3}{4}\Delta Y = \Delta I$$
$$\Delta Y(1 - 3/4) = \Delta I$$
$$\Delta Y = \frac{1}{1 - 3/4}\Delta I$$

and

$$\frac{\Delta Y}{\Delta I} = \frac{1}{1 - 3/4} = 4$$

If $\Delta I = 10$, $\Delta Y = 40$; the new equilibrium level of income and output will be 40 above the original level. Since $\Delta C = 3/4\ \Delta Y$ and since $\Delta Y = 40$, $\Delta C = 30$. The rise in income and output is divided between a rise in consumption of 30 and a rise in investment of 10.

Given any change in investment, ΔI, the change in income and output necessary to re-establish equilibrium is known as soon as the multiplier is known. The multiplier, in turn, is known as soon as the MPC is known. Thus, as we have just seen, the general expression for the multiplier is

$$\frac{\Delta Y}{\Delta I} = \frac{1}{1 - \text{MPC}}$$

In other words, the multiplier is the reciprocal of 1 minus the MPC, so that the larger the value of the MPC, the larger will be the value of the multiplier. This is clearly in accord with our intuitive notion that the rise in income induced by a given rise in investment will be larger the larger the proportion of that investment outlay which is respent—that is, the larger the MPC. Thus, for any value of ΔI, the greater the size of the multiplier, the greater will be the increase in the equilibrium level of income.

There is a second approach to the determination of the multiplier. To determine the equilibrium level of income and output, we earlier used $Y = C + I$ and also

$$S = I \qquad [1]$$

Whereas $Y = C + I$ focuses on the equality between aggregate output and aggregate spending, $S = I$ focuses on the equality between saving and investment. Once given the saving function and the investment function, the equation could be readily solved for the equilibrium level of income and output as follows:

$$S = I$$
$$-20 + \tfrac{1}{4}Y = 20$$
$$\tfrac{1}{4}Y = 40$$
$$Y = 160$$

The further development of the approach in terms of saving and investment is exactly parallel with that above in terms of consumption and investment. Assuming an upward shift in the investment function, the new equilibrium level of income and output is that at which

$$S + \Delta S = I + \Delta I \qquad [2]$$

Since $S = I$, we may subtract $S = I$ from Equation [2] which gives us

$$\Delta S = \Delta I \qquad [3]$$

The solution to this equation indicates the change in the level of income necessary to produce the new equilibrium level of income, given a specified change in the investment function.

The saving function, $S = S_a + sY$, indicates that saving rises or falls by an amount equal to the MPS or s (here 1/4) times the rise or fall in income. That is, $\Delta S = s\,\Delta Y$. Substituting in Equation [3], we have:

$$s\,\Delta Y = \Delta I$$

$$\Delta Y = \frac{1}{s}\,\Delta I$$

$$\frac{\Delta Y}{\Delta I} = \frac{1}{s}$$

If, as before, $\Delta I = 10$, then $\Delta Y = 40$, or the new equilibrium level of income and output will be 40 above the original level. Given any change in investment, ΔI, the change in income and output necessary to restore equilibrium is known as soon as the multiplier is known. The multiplier, in turn, is known as soon as the MPS is known. Therefore, we have as a second general expression for the multiplier:

$$\frac{\Delta Y}{\Delta I} = \frac{1}{\text{MPS}}$$

Since MPS $= 1 -$ MPC, 1/MPS as here derived is exactly equal to our earlier equation for the multiplier, $1/(1 - \text{MPC})$. Thus, the multiplier is the reciprocal of 1 minus the MPC or the reciprocal of the MPS.

SIMPLE INCOME DETERMINATION—A CONCLUDING NOTE

In this chapter we have examined the theory of income determination under some highly simplifying assumptions—hence the use of the adjective "simple." Some of these assumptions will be dropped in later chapters. Naturally, the more of them that are dropped, the more complicated the theory becomes and the closer it comes to describing the infinitely complex process by which income and output change in the real world.

There is no doubt, however, that the simple theory outlined in this chapter sheds considerable light on the aggregate economic process in the world about us; it tells us things that are not immediately apparent. One may properly argue that it takes no high level of economic sophistication to recognize that output will be increased in an economy with idle men and idle machines if either business persons or consumers or both step up their spending for goods and services. However, the completely unsophisticated will most likely expect output to increase by the amount of this initial increase in spending. But even the simple theory, merely by introducing the concepts of the consumption function, the marginal propensity to consume, and the multiplier, makes very clear the secondary consequences that produce an increase in income and output larger than the initial increase in investment or consumer spending.

Our simple theory not only shows the process by which an increase in autonomous spending will raise income and output by a multiple of that increase, but also indicates what determines the size of this multiple, admittedly under very restrictive assumptions. Although this simple theory cannot begin to explain the actual

fluctuations in the level of income and output in the real world, it can explain an important truth, a truth necessary but not sufficient to an understanding of these fluctuations. To understand this is to understand the essence of what the simple theory has to tell us.

chapter 6

\6

Government Spending and Taxation

In the preceding two chapters we set out the basic model of income determination for an economy composed of only consumers and business firms. In this chapter we expand that basic model to include the government as a third sector and work out the effects of government expenditures and taxation on the level of income and output. We will be dealing with aggregate spending composed of personal consumption, domestic investment, and government expenditures for final product and an aggregate flow of income that is now allocated not only to consumption and private saving but also, in part, to taxes.

In general, government can expand aggregate spending in any time period by increasing the amount it adds to the stream of private spending through its purchases of goods and services or by decreasing the amount it diverts from the stream of private spending through its net tax collections. By the same token, it can contract aggregate spending in any time period by decreasing the amount it adds to the stream of private spending through its purchases of goods and services or by increasing the amount it diverts from the stream of private spending through its net tax collections. The effect of government spending and taxation on aggregate spending thus depends, in the first instance, on how much government injects into the spending stream through its purchases and on how much it withdraws through net tax collections. Since the level of income and output depends on aggregate spending, government can clearly raise or lower the level of income and output through its policy with respect to spending and taxing.

This at least is a conclusion with which the vast majority of economists agree. There are a few extreme monetarists who seem to say that government is powerless to affect aggregate spending via its spending and taxing except to the extent that these involve a deficit that is financed by an increase in the money supply or a surplus that is used to reduce the money supply. They believe that government spending and taxing can affect aggregate spending only through the money supply changes that may accompany that spending and taxing. Although everyone now grants that the way a government deficit is financed or a surplus is disposed of significantly affects aggregate spending, most economists still hold that deficits and surpluses have a considerable effect on aggregate spending in and of themselves and apart from any change in the money supply that accompanies them. In this chapter, we study the way in which this takes place.

FISCAL POLICY

Government policy with respect to spending and taxing is known as its fiscal policy. In the years since the Great Depression, it has become generally accepted that the fiscal policy of the federal government should contribute to the attainment of certain economic goals. If the economy is operating at a level of income and output below that at which there is reasonably full utilization of its resources, the appropriate fiscal policy is an expansionary one. If, on the other hand, the economy is at a level of income and output at which there is not only full utilization of resources but also strong upward pressure on prices, the appropriate fiscal policy is a contractionary one. In other words, fiscal policy should operate in a countercyclical fashion, promoting the stabilization of economic activity at high levels of output and employment. Other goals of fiscal policy include rapid economic growth, greater equality in the distribution of income, and maximum economic "well-being." These goals overlap to a degree; some compete with each other, and some complement each other. In order to avoid the complexities of dealing with a diversity of goals, we will simply take the goal of fiscal policy to be the stabilization of economic activity at its full-employment level. Taking full employment to mean full utilization of all of the economy's resources, we may refer to the corresponding output as the economy's full-employment output.

Accepting this as the goal of fiscal policy, we would next want to devise the set of fiscal policy actions that would best enable us to achieve this objective under the conditions with which the economy is confronted at a particular time. There are two sets of basic fiscal policy alternatives. If the need is for an expansion of income, the fiscal policy alternatives are to increase government spending, decrease taxes, or both. If, on the other hand, due to inflationary pressures, there is a need to contract income, the fiscal policy alternatives are to decrease government spending, increase taxes, or both.

How does the government select the most effective alternative in any given situation? If there is a need for an expansion of income, what is the difference between increasing government purchases by a given dollar amount and decreasing tax collections by an equal amount? Can government produce any expansion in income by increasing its purchases by a given dollar amount if at the same time it increases its tax collections by an equal amount? Is there any difference between the expansionary effect of a given dollar increase in government purchases of goods and services and an equal increase in government transfer payments? To answer questions of this sort requires some understanding of the essential mechanics of fiscal policy.

To explain the mechanics of fiscal policy we will construct a series of three models, each of which is built on the models developed for the two-sector economy. In the first, only tax receipts, T, and government purchases, G, are added to the two-sector model; government transfer payments are in effect assumed to be zero. In the second model, government transfer payments are added. Both of these models assume that tax receipts are independent of the level of income—that they are "autonomous," to use the term adopted earlier to describe consumption and investment expenditures that are independent of the level of income. In the third model, the breakdown of government expenditures into purchases of goods and services and transfer payments is retained, but tax receipts are recognized as being, in part, dependent on the level of income. Since increases and decreases in government expenditures have expansionary or contractionary effects on the

level of income, it is found in this model, as is true in practice, that tax receipts depend in part on the level of government expenditures.[1]

Although, as we mentioned earlier, one accepted goal of fiscal policy is to promote full employment of the economy's resources, we will not attempt to show specifically in these three models how the level of government spending and taxation might be varied in order to accomplish this particular objective. The models simply assume certain amounts of government spending and taxation and indicate the expansionary or contractionary effect of each. The emphasis in this chapter will be on the pure mechanics of the relationships between government spending, taxation, and the level of income. However, in a final section of the chapter we will note briefly, in terms of a fiscal model, how the level of government spending and taxation might be adjusted to raise the level of output to the full-employment position. More will be said about full-employment fiscal policy in Chapter 24 of the book, which is devoted in part to the question of economic stabilization.

FIRST FISCAL MODEL—INCLUDING NET TAXES AND GOVERNMENT PURCHASES

In working out accounting identities in Chapter 2, the GNP identity for a three-sector economy was given as $C + S + T = \text{GNP} = C + I + G$. For NNP, here also designated by Y, the only change is that I is net rather than gross private investment and S is net rather than gross private saving. On both sides, the net figure is obtained from the gross figure by subtracting an amount equal to capital consumption allowances. With I and S now understood to refer to net amounts, we may write

$$C + S + T = Y = C + I + G$$

From this net national product identity follows the identity for saving and investment:

$$S + (T - G) = I$$

where $(T - G)$ equals public saving.

In the two-sector economy in which we assumed there were no undistributed corporate profits, disposable personal income was found to be equal to net national product. In the three-sector economy, however, with taxes taking a portion of the income flow generated by expenditures on net national product, disposable personal income is less than net national product by the amount of taxes. Letting Y represent net national product and Y_d disposable personal income, we now have

$$Y_d = Y - T$$

or

$$Y = Y_d + T$$

The consumption function for the two-sector economy, in which Y_d equaled Y, was $C = C_a + cY$. Now, with Y_d less than Y, the consumption function becomes $C = C_a + c(Y - T)$, or $C = C_a + cY_d$. With this consumption function, with investment assumed to be entirely autonomous, $I = I_a$, and with fixed amounts of government purchases and tax receipts assumed per time period, the equilibrium level of income is given by

$$Y = C_a + c(Y - T) + I + G$$

Expressed in terms of saving and investment, equilibrium will be found at that level of income and output at which planned saving plus taxes equals planned investment plus government purchases:

$$S + T = I + G$$

[1] For a more detailed development of these same models, see N. F. Keiser, *Macroeconomics, Fiscal Policy, and Economic Growth*, Wiley, 1964, Ch. 8, especially pp. 135–47. A more advanced and theoretical treatment is found in B. Hansen, *The Economic Theory of Fiscal Policy*, Harvard Univ. Press, 1958, and in A. Peacock and G. K. Shaw, *The Economic Theory of Fiscal Policy*, 2nd ed., Allen and Unwin, 1976.

For example, assume that $c = 3/4$, $C_a = 20$, and $I = 20$ and, for the moment, that T and G are both zero (which amounts to a two-sector economy). These are precisely the values assumed in the very first model of simple income determination in a two-sector economy, as illustrated in Figure 4-3. There we saw that the equilibrium level of income for the values given was 160. Now let us superimpose first a given level of government purchases and second a given level of taxes, and then trace the effects of each on the equilibrium level of income and output.

In Figure 6-1, the C and $C + I$ curves in Part A, the S and I curves in Part B, and the AD and AS curves in Part C correspond exactly to their counterparts in Figures 4-3 and 4-4. The equation for C is $C = C_a + c(Y - T)$, in which $C_a = 20$, $c = 3/4$, and, for the moment, $T = 0$. The equation for I is simply $I = I_a$ in which $I_a = 20$. Investment is completely autonomous or is the same at all levels of income. The equation for the saving function in Part B follows from the consumption function: $S = S_a + s(Y - T)$, in which $S_a = -C_a = -20$, and $s = 1 - c = 1/4$. The equilibrium level of income for these assumed values is, of course, 160 in Figure 6-1 as it was in Figures 4-3 and 4-4, and the aggregate demand curve for these assumed values is positioned at income of 160 in Figure 6-1 as it was in Figures 4-3 and 4-4. Now we insert G of 25 and assume that this amount is entirely deficit financed so that taxes remain at zero. This is shown in Part A of Figure 6-1 by simply adding the "autonomous" public spending, G of 25, to private spending, $C + I$, to produce the aggregate spending function, $C + I + G$, and, in Part B, by simply adding it again to investment spending, I, to produce the function labeled $I + G$. The new equilibrium level of income is now 260, 100 above what it was before aggregate spending was increased by the 25 of public spending. Correspondingly, in Part C this gives us at Y of 260 a new AD curve labeled AD_G in which the subscript indicates that this AD curve, unlike the one at Y of 160, includes government purchases of the amount shown in Parts A and B.

Why did the equilibrium level of income rise by 100 with the addition of government expenditures of only 25? With government now in the picture, the new equilibrium level of income must be that at which the amount of the income stream diverted from consumption into planned saving and tax payments is just equal to the amount added to the income stream in the form of investment and government expenditures. Since taxes are for the moment assumed to be zero, to restore equilibrium following the injection of 25 in government spending, income must rise to the level at which planned saving alone is equal to the sum of planned investment and government expenditures. Given an MPC of 3/4 and an MPS of 1/4, it is clear that only when income has risen by 100 will there be additional saving of 25 leaking out of the income stream per period to match the 25 injected into the income stream per period by government spending. Thus, in Figure 6-1, we see that $S = 45 = I + G$ at Y of 260. Once income has risen by 100, a new equilibrium is established.

The effect of adding a given amount of government expenditures, G, to the present model is essentially no different from the effect of an equal increase in C_a or I_a. Like C_a or I_a, G is here a part of autonomous spending. With $c = 3/4$, the ordinary multiplier is 4, and any increase in autonomous spending (e.g., G of 25) will, other things being unchanged, raise income by four times the increase in autonomous spending to restore the system to equilibrium. The equation for the equilibrium level of income may be rewritten to show the multiplier explicitly. That is,

$$Y = C_a + c(Y - T) + I + G$$

may be rewritten as

$$Y = \frac{1}{1 - c}(C_a - cT + I + G)$$

Solving for the values we have assumed for c, C_a, I, G, and T, we have

$$260 = 4[20 - \tfrac{3}{4}(0) + 20 + 25]$$

Now let us suppose that government abandons its policy of deficit spending and that it

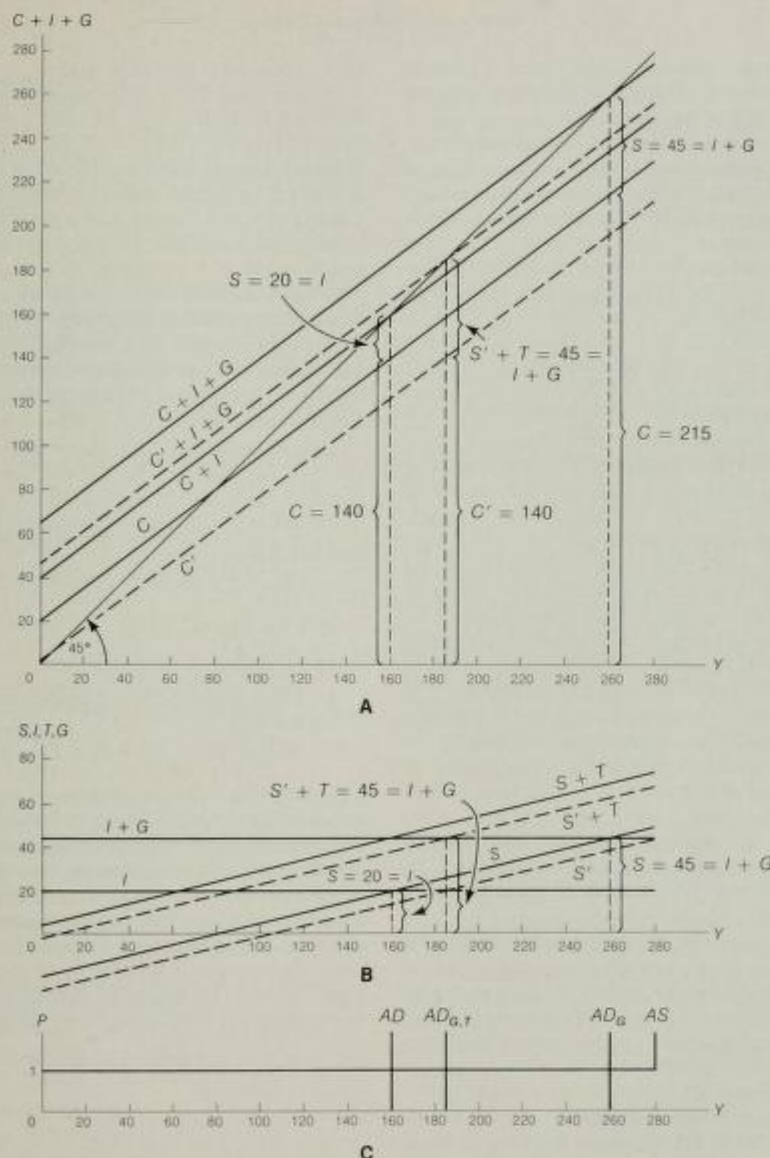


FIGURE 6-1

Equilibrium level of income, including government expenditures and taxation

collects in taxes 25 per time period to cover its expenditures of 25 per time period. With the imposition of this fixed amount of taxes, disposable income is 25 less than net national product at all levels of net national product. With an MPC of 3/4, it follows that taxes of 25 reduce consumption by 18.75 at each level of net national product. This is illustrated in Part A of Figure 6-1 by the new consumption function labeled C', which is positioned 18.75 below C. With an MPS of 1/4, taxes of 25 also reduce saving by 6.25 at each level of net national product. This is illustrated in Part B of Figure 6-1 by the new saving function labeled S', which lies 6.25 below S.

The new equilibrium level of income is now found in Part A by adding to C' the 20 of I and 25 of G to produce the new aggregate spending function, $C' + I + G$. This function intersects the guideline at Y of 185. The equilibrium level may also be found in Part B by adding to S' the 25 of T to produce $S' + T$, which intersects $I + G$ at Y of 185. In Part A, equilibrium is indicated at the level of Y at which aggregate spending equals aggregate output or at which $Y = C' + I + G$; in Part B, equilibrium is at the level of Y at which the aggregate diversions of income from consumption are offset by the aggregate of compensating expenditures, or the level of Y at which $S' + T = I + G$. This case, including G and T both equal to 25, gives us another aggregate demand curve, this one positioned at Y of 185 and labeled $AD_{G,T}$ to indicate that both G and T of the indicated amounts underlie this curve.

It is surprising at first glance to find that the injection by government of 25 into the income stream and the withdrawal in taxes of an exactly equal amount from the income stream still results in an expansion of the income stream from 160 to 185, an amount equal to the increase in the government budget. In the simplest terms, this result follows from the fact that the downward shift in the aggregate spending function due to T of 25 is less than the upward shift in that function due to G of 25. With the MPC of 3/4, T of 25 lowers the consumption function by 18.75 at each level of Y, or from C to C'. Since the upward shift of 25 resulting from G of 25 is 6.25 greater than the downward shift of 18.75 resulting from T of 25, the aggregate spending function with T and G both 25 is 6.25 above the aggregate spending function with T and G both zero. The net shift upward of 6.25 is subject to a multiplier of 4, which raises the equilibrium level of income by 25, from 160 to 185. As in previous models, this may also be explained in terms of saving and investment. Expressed in equation form, we have the following figures for the equilibrium positions, first with G and T both zero, and second with G and T both 25:

$$Y = C_a + c(Y - T) + I + G$$

$G = 0; T = 0$:

$$160 = 20 + \tfrac{3}{4}(160 - 0) + 20 + 0$$

$G = 25; T = 25$:

$$185 = 20 + \tfrac{3}{4}(185 - 25) + 20 + 25$$

$$S + T = I + G$$

$G = 0; T = 0$:

$$-20 + \tfrac{1}{4}(160 - 0) + 0 = 20 + 0$$

$G = 25; T = 25$:

$$-20 + \tfrac{1}{4}(185 - 25) + 25 = 20 + 25$$

Substitutions in either equation of larger or smaller amounts of G and T will, provided that G and T remain equal, result in a rise or fall in the equilibrium level of income equal to the amount of the increase or decrease in the size of the government budget. This result is known as the *balanced-budget theorem or unit-multiplier theorem*. We may see its meaning more clearly by again rewriting the aggregate spending equation to show the multiplier explicitly. That is,

$$Y = C_a + c(Y - T) + I + G$$

may be rewritten as

$$Y = \frac{1}{1 - c}(C_a - cT + I + G)$$

In this form, it may be seen that a change in any of the values within the parentheses will,

assuming that the other values within the parentheses remain unchanged, produce a change in income equal to the change in that value times the ordinary multiplier, $1/(1 - c)$. Substituting the values from the previous example, we have

$$185 = \frac{1}{1 - 3/4}[20 - \tfrac{3}{4}(25) + 20 + 25]$$

If now, for example, we assume a change in I, the other values remaining unchanged, the new equilibrium level of Y is equal to the original level of Y plus the change in Y:

$$Y + \Delta Y = \frac{1}{1 - c}(C_a - cT + I + G) + \frac{1}{1 - c}\Delta I$$

Subtracting Y from both sides, there remains

$$\Delta Y = \frac{1}{1 - c}\Delta I = \frac{\Delta I}{1 - c}$$

In the same way, we will find that

$$\Delta Y = \frac{1}{1 - c}\Delta C_a = \frac{\Delta C_a}{1 - c}$$

and

$$\Delta Y = \frac{1}{1 - c}\Delta G = \frac{\Delta G}{1 - c}$$

Therefore, for equal sized changes in I, C_a, and G, we have

$$\Delta Y = \frac{\Delta I}{1 - c} = \frac{\Delta C_a}{1 - c} = \frac{\Delta G}{1 - c}$$

However, a change in T of the same size as the change in I, C_a, or G will produce a smaller change in Y and one that is in the opposite direction, since

$$\Delta Y = \frac{-c\,\Delta T}{1 - c}$$

We saw in the example above that the equilibrium level of income with $G = 0$ and $T = 0$ was 160. Then with $G = 25$ and $T = 0$, the equilibrium level rose from 160 to 260 or $\Delta Y = 100$. That is,

$$\Delta Y = \frac{\Delta G}{1 - c} = \frac{25}{1 - 3/4} = 100$$

However, with both $T = 25$ and $G = 25$ in the model, we found the new equilibrium level to be 185. The addition of $T = 25$ to the model pulled income down, since

$$\Delta Y = \frac{-c\,\Delta T}{1 - c} = \frac{-\tfrac{3}{4}(25)}{1 - 3/4} = -75$$

Adding G of 25 raised the equilibrium level by 100, but adding T of 25 pulled the equilibrium level down by 75. Putting both together,

$$\Delta Y = \frac{\Delta G}{1 - c} + \frac{-c\,\Delta T}{1 - c} = 100 - 75 = 25$$

The change in Y is 25, as noted earlier.

The contractionary effect of an increase in taxes is thus less than the expansionary effect of an equal increase in government spending for goods and services. A rise in G is, in its entirety, an addition to aggregate spending, but a rise in T is not, in its entirety, a decrease in aggregate spending. Some part of the rise in T is absorbed by a decrease in S, and only the remainder by a decrease in C, or in aggregate spending.

The difference in the impact of ΔG and ΔT on the level of Y may also be seen by comparing the multipliers that apply to each. The government purchases multiplier is the same as the multiplier applicable to a change in autonomous consumption or investment spending. Thus, for ΔG,

$$\Delta Y = \frac{\Delta G}{1 - c}$$

or

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1 - c}$$

For ΔT, however, we derive what is called the *tax multiplier*, which is smaller than the government purchases multiplier. Whereas a change in government purchases leads to a change in autonomous spending that exactly equals the change in government purchases, a change in taxes leads to a change in autonomous spending that is only c times the change in taxes. Thus, for ΔT,

$$\Delta Y = \frac{-c\,\Delta T}{1 - c}$$

or

$$\frac{\Delta Y}{\Delta T} = \frac{-c}{1 - c}$$

If, as earlier, we assume c to be $3/4$, a rise

in government purchases is subject to a multiplier of 4, but a rise in taxes is subject to a multiplier of -3. In other words, an additional dollar of G will raise Y by \$4, and an additional dollar of T will reduce Y by \$3, leaving a \$1 rise in Y as the net effect of a rise of \$1 in both G and T.

Regardless of the value of c, the government purchases multiplier, $\Delta Y/\Delta G$, will always be 1 greater than the tax multiplier, $\Delta Y/\Delta T$. This may be shown by combining the separate multiplier expressions for ΔG and ΔT. Thus,

$$\frac{\Delta Y}{\Delta G} + \frac{\Delta Y}{\Delta T} = \frac{1}{1-c} + \frac{-c}{1-c} = \frac{1-c}{1-c} = 1$$

Since $\Delta Y/\Delta G$ is always positive and $\Delta Y/\Delta T$ always negative, the sum of the two will always be unity, whatever the value of c.

A dramatic implication for fiscal policy seems to follow from the unit-multiplier theorem. If the level of the economy's output is below full employment, it would seem that government can raise the level to full employment by an appropriate expansion in the size of its budget, covering every dollar of additional expenditures with a dollar of additional taxes. The desired rise in income and output may thus be achieved by means of a fiscal policy that does not resort to deficit financing, with its real or alleged "evils." However, as we have noted a number of times before, these crude mechanical models are in each case subject to numerous qualifications that complicate the solution. The road to full employment is certainly not so simple as is suggested by the crude unit-multiplier theorem. Although it is not certain that a rise in the size of the budget, with taxes and expenditures both up by, say, \$5 billion, will raise the level of income by \$5 billion, it certainly will not be neutral in its effect on the level of income. To analyze the expansionary effects of a balanced budget properly involves more than the unit-multiplier theorem, but the mechanical model of that theorem as developed here at least dispels the notion that a balanced budget is fiscally neutral, as was once thought to be the case.[2]

SECOND FISCAL MODEL—INCLUDING GROSS TAXES, GOVERNMENT PURCHASES, AND TRANSFER PAYMENTS

The first model emphasized the effects on income of changes in the net tax receipts of government, T, and government purchases of goods and services, G. Now let us introduce a simple modification that brings out the essential difference between the effects on income of changes in government purchases and of changes in government transfer payments.

Net tax receipts, T, are equal to gross tax receipts minus government transfer payments or $T_g - R$.[3] Expressing this as $T = T_g - R$ underscores the fact that R is really negative taxes, in effect an amount of gross tax receipts that is returned to individuals through government transfer payments.[4] Substituting $T_g - R$ for T, the fundamental identity for net national product now becomes

$$C + S + T_g - R = Y = C + I + G$$

[2]The literature on the unit-multiplier theorem is summarized in W. J. Baumol and M. H. Peston, "More on the Multiplier Effects of a Balanced Budget," in *American Economic Review*, March 1955, pp. 140–48. The more important studies in this extensive literature are also listed in R. A. Musgrave, *The Theory of Public Finance*, McGraw-Hill, 1959, p. 430.

[3]More exactly, one must subtract the sum of transfer payments, government net interest payments, and subsidies less current surplus of government enterprises, as explained in Chapter 2, p. 28. R will be referred to here simply as transfer payments but technically it includes the other transfer-type expenditures noted.

[4]With a balanced budget, $G = T$ and $G + R = T_g$. By showing R as a deduction from gross tax receipts, we have $G = T_g - R$. If there is a deficit or surplus, $G > T$ or $G < T$; the size of the deficit or surplus, of course, remains unchanged in the equation as restated.

Disposable personal income, in turn, becomes

$$Y_d = Y - T_g + R$$

The consumption function becomes

$$C = C_a + c(Y - T_g + R)$$

And the equilibrium level of income is given by

$$Y = C_a + c(Y - T_g + R) + I + G$$

which may be rewritten as

$$Y = \frac{1}{1-c}(C_a - cT_g + cR + I + G)$$

This equation for the equilibrium level of income is identical to that for the first model, except that $T_g - R$ is now substituted for T. And the interpretation of this equation as rewritten is identical to that of its counterpart in the first model—a change in any of the values within the parentheses, on the assumption that all other values within the parentheses remain unchanged, will produce a change in income equal to the change in that value times the ordinary multiplier.

From the equation it is clear that the effect on Y of an increase in R will be less than the effect of an equal increase in G, as long as the MPC, or c, is less than 1. That is,

$$\frac{1}{1-c}\Delta G > \frac{1}{1-c}c\Delta R$$

where ΔG equals ΔR.

The reason for this difference is that, in the first instance, all of any increase in G is an addition to aggregate spending, whereas only part of any increase in R becomes an addition to aggregate spending. ΔG affects aggregate spending directly, but ΔR affects aggregate spending only indirectly through its effect on disposable income. Assuming that there is no change in tax receipts, ΔR increases disposable income directly by the full amount of ΔR. The consumption function indicates, however, that not all of any increase in disposable income will be devoted to consumer spending; some portion will be devoted to personal saving. *In other words, at the first step, some portion of government transfer payments will fail to appear as spending for goods and services, but all government purchases will appear as spending for goods and services.* Thus, in the case of government purchases, the full increase in government spending is subject to the ordinary multiplier, but in the case of government transfers only the part that is not diverted into saving is subject to the ordinary multiplier. These amounts may be designated ΔG and $c\Delta R$, respectively, as in the equation given above for the equilibrium level of income.

Instead of showing that all of ΔG and only part of ΔR are subject to the ordinary multiplier, we may express the same thing differently in terms of the government purchases multiplier and the government transfers multiplier. For government purchases, we developed the following multiplier:

$$\Delta Y = \frac{1}{1-c}\Delta G$$

or

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1-c}$$

For government transfers, we derive the multiplier to which the total of any change in transfers is subject as follows:

$$\Delta Y = \frac{1}{1-c}c\,\Delta R$$

or

$$\frac{\Delta Y}{\Delta R} = \frac{c}{1-c}$$

Regardless of the value of c, the government transfers multiplier is 1 less than the government purchases multiplier. Apart from the change in sign, it is the same as the tax multiplier. This equality between the tax multiplier and the government transfers multiplier has several important implications that will be noted below.

First, for a numerical example, let us compare the effects of ΔG of 5 with ΔR of 5, assuming that c equals 3/4:

$$\Delta Y = \frac{1}{1-c}\Delta G = 20$$

and

$$\Delta Y = \frac{1}{1-c} c\,\Delta R = 15$$

Expressed in terms of the government purchases multiplier and the government transfers multiplier, this means that ΔG is subject to a multiplier of 4 and ΔR to a multiplier of 3.

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1-c} = 4$$

and

$$\frac{\Delta Y}{\Delta R} = \frac{c}{1-c} = 3$$

Since ΔG exerts a greater expansionary effect on Y than ΔR exerts when T_g remains constant, one would also expect the expansionary effect of ΔG on Y to exceed that of ΔR when ΔT_g matches the increase in government expenditures. The balanced-budget theorem showed that an equal change in both G and T (for which we may now substitute T_g) would produce a change in Y equal to the change in the size of the budget. For example, with ΔG and ΔT_g both 5 and with c of 3/4, we found the following:

$$\Delta Y = \frac{\Delta G}{1-c} + \frac{-c\,\Delta T_g}{1-c} = 20 - 15 = 5$$

Now, however, if the tax-financed increase in government expenditures is an increase in R rather than in G, we would have a different result.

$$\Delta Y = \frac{c\,\Delta R}{1-c} + \frac{-c\,\Delta T_g}{1-c} = 15 - 15 = 0$$

Unlike an increase in G, the expansionary effect of an increase in R is fully offset by the contractionary effect of an equal increase in T_g. In short, the expansionary effect suggested by the balanced-budget theorem applies not to tax-financed changes in R but only to tax-financed changes in G.

In practice, the differences between the expansionary effects of an increase in purchases as against those of an increase in transfer payments are not likely to be precisely those indicated by these crude models.[5] Despite the need for qualifications, the models still permit us to reach some tentative conclusions concerning the probable results of alternative fiscal policies. These conclusions may be summarized as follows. Assuming that government may, at its discretion, alter one or more of the variables according to plan, it may incur a deficit by (1) reducing tax receipts, (2) increasing purchases, (3) increasing transfer payments, or (4) any combination of these changes. In making a deficit of a given amount, an increase in purchases will have a greater expansionary effect than either a reduction in tax receipts or an increase in transfer payments. The expansionary effect of a deficit of given amount incurred by reducing tax receipts or raising transfer expenditures will be of the same order. Assuming that additional expenditures are to be covered by additional taxes, government can produce an expansionary effect only by increasing its purchases of goods and services. A reduction in its purchases accompanied by an equal reduction in tax receipts may be expected to be contractionary. In short, if government policy is to avoid a change in the size of the deficit or surplus, increasing or decreasing transfer payments and tax receipts in like amount will not directly affect the level of income.

THIRD FISCAL MODEL—INCLUDING GROSS TAX RECEIPTS AS A FUNCTION OF INCOME, GOVERNMENT PURCHASES, AND TRANSFER PAYMENTS

In the previous model, we developed the following equation for the equilibrium level of income, from which we may find, for any given value of c, the effect that a change in any one element within parentheses will have on Y if all the other elements remain constant:

[5] Just one of a number of reasons is that the crude models assume that the MPC is the same for all persons. However,

$$Y = \frac{1}{1 - c}(C_a - cT_g + cR + I + G)$$

In reality, of course, a change in any one element is bound to affect the others through its effect on the level of income. A change in spending, whether directly through a change in autonomous consumption, C_a, investment, I, or government purchases, G, or indirectly through a change in government transfer payments, R, or tax receipts, T_g, will change the level of income; this change may call forth changes in other components of spending. The response of consumption spending to a change in income is already built into the model because consumption is assumed to be equal to $C_a + cY_d$. The response of investment spending could also be built into the model by replacing the investment function in which investment is completely autonomous with one that makes investment depend in one way or another on the level of income or on changes in the level of income.[6] The response of government spending to changes in the level of income could also be built into the model, except that there is no simple, meaningful relationship between government spending, apart from transfer payments, and the level of income. To keep our discussion of this third model comparatively simple, we will continue to treat investment spending and government spending as completely autonomous.

One modification may easily be made that will bring our third model closer to reality, and that is to allow for the fact that any change in income will typically affect tax receipts. In the previous model we assumed that tax receipts would remain constant in the face of a change in income. In practice this would happen only if government took just the offsetting action necessary to prevent the change in tax receipts that otherwise would follow automatically from a change in income. In recent years, well over half the combined tax receipts of federal, state, and local governments have been from personal and corporate income taxes, and the revenue produced by these taxes, far from being independent of income, varies more than proportionally with changes in income. Other mainstays in the tax structure, such as sales and excise tax receipts, are also related to income but less closely than income tax receipts. Given our present tax and rate structure, to prevent the changes in tax receipts that would automatically accompany a change in income would call for appropriate changes in the tax and/or rate structure. Now, assuming, as is found in practice, that tax receipts do vary with changes in income, we may treat tax receipts roughly as a linear function of income. This gives us the following tax function:

$$T_g = T_a + tY$$

A tax function of this type is depicted by the T_g line in Figure 6-2. Notice that this function is of the same type as the consumption function, $C = C_a + cY$. Accordingly, it is interpreted in the same way. Just as c represents the MPC, so t stands for the MPT or the *marginal propensity to tax*. It indicates the marginal tax rate, $\Delta T_g/\Delta Y$, or the fraction of any change in income that will be diverted from income receivers to government with a given tax structure and tax rates. It is comparable to the single rate in a proportional income tax. Graphically, it is the slope of the T_g function, drawn in Figure 6-2 as 1/5. This tells us that for every change in Y of \$5, there will be a change in T_g of \$1. T_a, which is analogous to C_a, represents the amount of "autonomous" tax receipts, or the amount of tax receipts that is independent of the level of

---

the MPC of persons suffering a reduction in disposable income through additional tax payments may well be different from that of persons enjoying an increase in disposable income through receipt of transfer payments. The decrease in consumption of taxpayers will, therefore, not necessarily be the same as the increase in consumption of the beneficiaries of the transfer payments. The two sets of persons need not be the same. For example, unemployment compensation is financed through business taxes whose incidence may be on the owners or the customers of business or on both, but the beneficiaries of the compensation checks are the unemployed.

[6] Several functions of this kind are developed in Chapter 11.

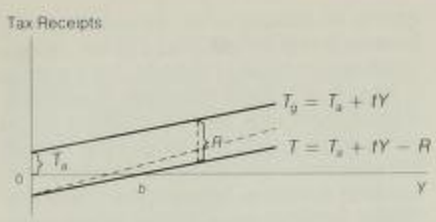


FIGURE 6-2
Gross and net tax functions

income. Although in practice income could not fall to zero, T_a indicates the amount of tax receipts at the theoretical zero level of income, which is depicted graphically by the intercept of the T_g function with the vertical axis.

The T_g function of Figure 6-2 shows T_a as a positive amount. Assuming that R remains the same at all levels of income and expressing the tax function in net terms, the net tax function, labeled T in the figure, would lie below the gross tax function by the amount of R. As before, R appears as negative tax receipts or as a subtraction from gross tax receipts. The net tax function, $T = T_a + tY - R$, crosses the vertical axis below zero, indicating that at very low levels of income, the amount of transfer payments would exceed the amount of gross tax receipts. Net tax receipts are shown to be negative at all income levels below $0b$. If we go a step further and recognize that transfer payments vary inversely with the level of income, the net tax function corresponding to the gross tax function given in Figure 6-2 would be one like the dashed line lying above T. This reflects the fact that as income rises, unemployment compensation and other expenditures under general relief programs decrease. This results in a narrowing of the spread between the gross tax function and the net tax function as income rises and a widening of the spread as income falls. For the sake of simplicity, however, we will assume that R is independent of Y, so that $T_g = T_a + tY$ and $T = T_a + tY - R$.

Substituting the tax function, $T_g = T_a + tY$, for T_g, we now have

$$C + S + T_a + tY - R = Y = C + I + G$$

Disposable personal income becomes

$$Y_d = Y - (T_a + tY) + R$$

or

$$Y_d = Y - T_a - tY + R$$

The consumption function becomes

$$C = C_a + c(Y - T_a - tY + R)$$

Retaining our assumption that investment and government expenditures are entirely autonomous, the equilibrium level of income is given by

$$Y = C_a + c(Y - T_a - tY + R) + I + G$$

or

$$Y = C_a + cY - cT_a - ctY + cR + I + G$$

which may be rewritten as

$$Y = \frac{1}{1 - c(1 - t)}(C_a - cT_a + cR + I + G)$$

With tax receipts assumed to be independent of the level of income, the equation for the second model had a multiplier of $1/(1 - c)$. Recognizing that tax receipts are dependent on the level of income, the multiplier is reduced to $1/[1 - c(1 - t)]$. Assuming that $c = 3/4$ and that $t = 1/5$, we now have the smaller multiplier 2.5, instead of the previous multiplier 4. If the marginal propensity to tax, t, is greater than zero, the present model will always yield a smaller multiplier than the previous model.

Thus, with $c = 3/4$ and $t = 1/5$, for a change in G of 10 we now have

$$\Delta Y = \frac{1}{1 - c(1 - t)}\Delta G$$
$$= \frac{1}{1 - (3/4)(1 - 1/5)}(10) = 25$$

instead of our earlier result,

$$Y = \frac{1}{1 - c}\Delta G = \frac{1}{1 - 3/4}(10) = 40$$

In the period in which G increases by 10, income increases by 10. If $t = 0$, disposable income also rises by 10, so that consumers

have available the full increase in income to spend or save. If we now assume that $t = 1/5$, disposable income does not rise by the amount of the increase in income. Instead, 1/5 of the income increase of 10 is diverted to the government in the form of tax payments, leaving only 8 as the increase in disposable income. With $c = 3/4$, instead of induced consumption of 7.5, or 3/4 of 10, as in the previous model, we now have induced consumption of 6, or $3/4 \times 4/5 \times 10$. Thus, the marginal propensity to tax operates as a drag on consumption, reducing it to 6 from the 7.5 it would be if t equaled zero. In each of the subsequent periods required for the multiplier process to work itself out, 1/5 of the income generated in each period is also diverted to government in the form of tax payments. The final result is a new equilibrium level of income with ΔY of 25, made up of ΔG of 10 and ΔC of 15.

TABLE 6-1
Marginal propensity to tax and the multiplier process

A: $t = 1/5, c = 3/4$					
PERIOD	ΔY	ΔG	ΔT_g	ΔY_d	ΔC
t	0	0	0	0	0
$t + 1$	10.0	10	2.0	8.0	0
$t + 2$	16.0	10	3.2	12.8	6.0
$t + 3$	19.6	10	3.9	15.7	9.6
$t + 4$	21.8	10	4.4	17.4	11.8
$t + 5$	23.0	10	4.6	18.4	13.0
. . .	. . .	. . .	. . .	. . .	. . .
$t + n$	25.0	10	5.0	20.0	15.0

B: $t = 0, c = 3/4$					
PERIOD	ΔY	ΔG	ΔT_g	ΔY_d	ΔC
t	0	0	0	0	0
$t + 1$	10.0	10	0	10.0	0
$t + 2$	17.5	10	0	17.5	7.5
$t + 3$	23.1	10	0	23.1	13.1
$t + 4$	27.3	10	0	27.3	17.3
$t + 5$	30.5	10	0	30.5	20.5
. . .	. . .	. . .	. . .	. . .	. . .
$t + n$	40.0	10	0	40.0	30.0

NOTE: Figures are rounded to nearest tenth.

This process is described in detail in Table 6-1. Part A assumes $t = 1/5$ and traces the multiplier process given by the multiplier, $1/[1 - c(1 - t)]$. Part B assumes that tax receipts are independent of income, $t = 0$, and traces the familiar multiplier process given by the ordinary multiplier, $1/(1 - c)$. In both parts, there is an original equilibrium in period t, which is upset in period $t + 1$ by a permanent increase of 10 in government purchases. For use in this model, a lagged consumption function is assumed, so a change in disposable income in any one period does not lead to a change in consumption spending until the following period.

In Part A, ΔG of 10 in period $t + 1$ generates ΔY of 10 but ΔY_d of only 8, since 1/5 of ΔY is diverted to government as ΔT_g. In period $t + 2$, ΔY is 16, which consists of ΔG of 10 and ΔC of 6, the latter being 3/4 of ΔY_d of period $t + 1$. Of ΔY of 16, 1/5 (or 3.2) is diverted to government, leaving ΔY_d of 12.8. In period $t + 3$, ΔY is then the sum of ΔG of 10 and ΔC of 9.6. Income gradually rises by period $t + n$ to a new equilibrium 25 above the original equilibrium. In Part B, the process is the same, except that no portion of ΔY is diverted into tax payments. Without additional taxes to be paid on the additional income, disposable income is greater, consumption spending is greater, and the level of income rises faster and farther than in Part A. In this situation, income rises to a new equilibrium level 40 above the old. Similarly, since consumption is not subject to the drag of additional taxes, induced consumption rises by 30 in Part B as compared with only 15 in Part A.

Although it is clear that a positive marginal propensity to tax means a smaller expansion of income for any given increase in government spending than a zero marginal propensity to tax, it also means that the deficit created by any increase in government spending will be less than the amount of the increase in government

spending. For example, in Part A the deficit created by ΔG of 10 is only 5 per time period, once the income equilibrium level has been reached—that is, $\Delta G - \Delta T_g = 5$ in period $t + n$. In contrast, we see in Part B, with $t = 0$, that ΔG of 10 creates a deficit of 10 in each period. Part A is more relevant, since in the real world the MPT is positive. This suggests that normally an increase in government expenditures, with no change in tax rates or tax structure, will lead to a less than equal increase in the deficit.

This point is sometimes missed by persons who overlook the expansion of the income flow produced by the increase in government spending. With a positive MPT, some part of the expanded income flow automatically becomes additional tax receipts of government and as such becomes a check to what otherwise would be a change in the deficit equal to the increase in government spending. As we can see in the present model, the greater the MPC, the greater will be this check to the size of the deficit accompanying any increase in government spending. If, for example, in Part A we substitute an MPC of 9/10 for the MPC of 3/4 and retain the same MPT of 1/5, the multiplier, $1/[1 - c(1 - t)]$ becomes 3.57. Then ΔG of 10 would produce a new equilibrium in period $t + n$ with ΔY of 35.7. With ΔY of 35.7, ΔT_g would be 7.14 in period $t + n$ and the deficit would be $10.00 - 7.14$ or only 2.86 in each period. Although an MPC of such size is unrealistic, in the present model an MPC that approaches 1 will produce a multiplier that approaches 5, given the same MPT of 1/5. In this case, ΔG of 10 would produce a new equilibrium in period $t + n$ with ΔY of 50. The result would then be a "balanced budget" because ΔT_g would be 10 in period $t + n$ or equal to ΔG. The increase in the equilibrium level of income induced by the increase in government spending would raise tax receipts by the same amount as the increase in government spending.

FISCAL MODELS AND THE FULL-EMPLOYMENT LEVEL OF INCOME

The several fiscal models developed to this point have shown how changes in G, R, and T_g, individually and in combination, affect the level of Y. We can summarize the implications of these simple fiscal models by demonstrating in terms of such a model how alternative fiscal policies may be used to lift an economy to its full-employment level of income and output.

Let us begin with the most recently developed equation for the equilibrium level of income.

$$Y = C_a + c(Y - T_a - tY + R) + I + G$$

or

$$Y = \frac{1}{1 - c(1 - t)}(C_a - cT_a + cR + I + G)$$

We will assume that the following values apply:

$$C = 15 + \tfrac{3}{4}Y_d \quad R = 10.66 \quad T_g = 4 + \tfrac{1}{5}Y$$
$$I = 32 \qquad G = 50$$

Substituting in the equation for the equilibrium level of income, we have

$$Y = 15 + \tfrac{3}{4}(Y - 4 - \tfrac{1}{5}Y + 10.66) + 32 + 50$$
$$Y = 255$$

with $C = 173$, $I = 32$, and $G = 50$.

The aggregate supply function adopted in this and the preceding chapters indicates that the full employment level of Y is 280.[7] The model suggests several possible routes by

[7]With the equilibrium level of Y at 280, a perfectly inelastic AD curve would be positioned at Y of 280 in Part C of Figure 6-1 and would thus coincide with the perfectly inelastic portion of the AS curve. In this special case, the equilibrium price level becomes indeterminate because $AD = AS$ at all possible price levels. However, this does not conflict with the establishment of the equilibrium output at Y of 280. The problem of an indeterminate price level at the full employment level of Y will be resolved by the analysis of Part 4 which yields a downward sloping AD curve. As may be apparent, such a curve appropriately

which this output level may be attained. An increase in government expenditures for goods and services is one route. To determine the needed increase in government expenditures, we insert Y of 280 into the equation and solve for the necessary G.

$$280 = 15 + \tfrac{3}{4}[280 - 4 - \tfrac{1}{5}(280) + 10.66] + 32 + G$$
$$G = 60$$

This could be found more simply and directly from the second equation given above, which shows the effective multiplier to be 2.5. To produce the required increase in Y of 25 would then call for an increase in G of 10 or from G of 50 to G of 60.

With Y at 255, T_g is 55 and $G + R$ is 60.66, indicating that there is a deficit of 5.66. If G rises by 10, resulting in a rise in Y of 25, T_g becomes 60, and the deficit increases to 10.66. Despite the increase in government expenditures of 10, the deficit increases by only 5; additional tax receipts of 5 are provided by the rise in income of 25.

A second route to full employment would be through tax reduction. Since tax receipts in the model depend on the level of income, tax reduction may take the form of a reduction in the marginal tax rate, t. To determine how far to reduce the tax rate, t, other things being equal, in order to produce aggregate spending of 280 and so income and output of 280, we insert Y of 280 into the equation and solve for the necessary value of t.

$$280 = 15 + \tfrac{3}{4}(280 - 4 - 280t + 10.66) + 32 + 50$$
$$t = 0.15$$

A reduction in the marginal tax rate from 0.20 to 0.15 (or more precisely to 0.1523) will raise income by the required amount. Or, other things being equal, this is the tax rate that will produce the level of disposable income at which, given the marginal propensity to consume, consumption spending will, when added to investment and government spending, yield the required level of aggregate spending of 280 and so the required income level of 280.

As before, the required tax rate may be found more simply and directly from the alternative equation that originally showed the effective multiplier to be 2.5. That is,

$$\frac{255}{102} = \frac{1}{1 - \tfrac{3}{4}(1 - 1/5)} = 2.5$$

For Y of 280, the effective multiplier is thus 280/102 or 2.75. To achieve this desired multiplier, t must be 0.15, as can be determined by solving the following for t:

$$\frac{280}{102} = \frac{1}{1 - \tfrac{3}{4}(1 - t)} = 2.75$$

As we mentioned earlier, the expansionary effect of a deficit incurred by way of tax reduction is less than that of an equal deficit incurred by way of increased government purchases. The same result is found here, except that we must now compare the size of the deficit required to produce a specific expansion in income by way of increased government purchases with that required to produce it by way of a decreased tax rate. Full employment was reached through increased purchases with an increase in the deficit from 5.66 to 10.66 per period. It is reached through a decreased tax rate with an increase in the deficit from 5.66 to 14.66 per period. In the present example, it takes an increase in the deficit of 9 per period via tax reduction to produce the expansion of income of 25 secured by an increase in the deficit of 5 per period via increased government purchases.

A third route to full employment would raise the income level by the required amount without any deficit financing. Since the expansionary effect of a dollar of government purchases is greater than the contractionary effect of a dollar of taxes, some amount of increase in government purchases and in taxes will raise

---

positioned will intersect a perfectly inelastic AS curve and thus indicate at that intersection an equilibrium rather than an indeterminate price level for the full employment level of Y.

income to the full-employment level and produce a balanced budget at the same time. To determine the level of G and T (net taxes) at which $G = T$ with $Y = 280$, we start off with:

$$280 = 15 + \tfrac{3}{4}(280 - 4 - 280t + 10.66) + 32 + G$$

Given that T (net taxes) at Y of 280 equals $4 + 280t - 10.66$, we substitute G for this expression and solve the equation for G:

$$280 = 15 + \tfrac{3}{4}(280 - G) + 32 + G$$
$$280 = 257 + \tfrac{1}{4}G$$
$$G = 92$$

Therefore, we know that T (net taxes) must equal 92 to produce a balanced budget with Y of 280. Corresponding to net tax receipts, T, of 92 are gross tax receipts, T_g, of 102.66. The required marginal tax rate, t, is therefore found from the tax function $102.66 = 4 + 280t$, or $t = 0.35$ (or more precisely 0.3523). The final solution is then

$$280 = 15 + 3/4[280 - 4 - 0.35(280) + 10.66] + 32 + 92$$
$$280 = 156 + 32 + 92$$

in which $C = 156$, $I = 32$, and $G = 92$.

In comparison with the first route, this one calls for a much greater expansion in government purchases. In comparison with the second route, it calls for an increase in the marginal tax rate. Although it manages to raise the income level without deficit financing, it does so only by enlarging the share of the flow of goods and services and the share of the flow of income that is absorbed by government. A greater fraction of what is now a greater flow of final goods and services is made up of governmentally provided goods and services. A greater share of what is now a greater flow of income is diverted into taxes, and a smaller share is left for private consumption and saving.

The choice of which of these three fiscal policies to use in a given situation depends on what set of side-effects is regarded as the least harmful. Government must evaluate the relative merits and demerits of spending, taxing, and changing the size of the public debt. In practice, the first and second routes or some combination of the two have received the most attention. Although both involve deficits and therefore some growth in the public debt, taxing does not enlarge the scope of government activities at all, and government spending, while enlarging the scope of government activities, produces the desired effect on the income level without the relatively much greater enlargement of government activities necessitated by a fiscal policy with the budget-balancing restraint.

A CONCLUDING NOTE

No matter how elaborate fiscal models may be, they are all built essentially along the lines of the simple models discussed in this chapter. The chief purpose of all such models is to shed light on the way in which government taxing and spending affect aggregate demand and thus the level of income and output. However, as with previous models, it should be emphasized that the conclusions suggested by the succession of models in this chapter are based on the assumption of "other things being equal." Yet other things may change as a direct result of the government's taxing–borrowing–spending process, and to the extent that they do, the conclusions suggested by the models must be modified. Whatever these other changes may be, they may be examined through their effects on the consumption and investment functions.

In the case of the consumption function, the fiscal process may alter the slope or level (intercept with the vertical axis) of the function itself.

The expansionary effect of a tax-financed increase in government purchases will be greater than is shown in the models if it leads to a rise in the MPC or in the level of the consumption function. Conversely, it will be smaller than is shown in the models if it leads to a fall in the MPC or in the level of the consumption function. One way the government's fiscal actions might produce such results is through their effect on the distribution of income. For example, to the extent that the MPC of the "poor" is greater than that of the "rich," government expenditures financed through a progressive income tax will raise the aggregate MPC and have a greater expansionary effect on the income level than if the aggregate MPC remains unchanged. If the same government expenditures were financed instead by sales taxation, they might conceivably lower the aggregate MPC and thus have a smaller expansionary effect on the income level than if the aggregate MPC had remained unchanged.

Even more important are the possible effects of the government taxing–spending process on the investment function. A tax-financed increase in government expenditures covered by higher corporate or progressive personal income tax rates might have an adverse effect on the willingness of investors to spend on structures and durable equipment. The resultant downward shift in the investment function could offset the expansionary effect of the rise in government spending. Even the case in which the increased government expenditures are deficit financed is not without possible adverse effects on the investment function. The current increase in the national debt may generate fears of worsening inflation, of a coming rise in taxes, or even of a deterioration in the government's credit standing. The fact that fears of the possible financial collapse of government are entirely unfounded does not prevent them from arising, and it does not prevent them from adversely affecting the investment function. Furthermore, with a debt-financed increase in government expenditures, there is also a possibility of an adverse effect on investment through a rise in the rate of interest. If an increase in government borrowing is added to the existing private demand for loanable funds, the interest rate may be expected to rise. The monetarists make much of this, arguing that the higher rate of interest will "crowd out" private investment spending of an amount equal to the debt-financed increase in government spending. In this event, the increase in government spending does not on balance have an expansionary effect on income. If the deficit is financed by an increase in the money supply, this offsetting may not occur. The central bank has the power to expand the money supply by the indicated amount, but often the central bank, rightly or wrongly fearful of inflation, will not accommodate the increased demand for funds. In this event, the interest rate will tend to rise as it does with an increase in government borrowing from the public. Again, according to the "crowding out" argument the resulting decrease in investment spending will offset the otherwise expansionary effect of the increase in government spending.

These are simply a few of the numerous qualifications that must be recognized when one works with the simple models discussed in this chapter. The actual results of any change in the government's taxing–borrowing–spending program cannot be predicted with a high degree of accuracy. Yet, to the extent that a general statement is possible, most will agree that an increase in government spending or a reduction in taxes will encourage both consumption and investment and thus produce a rise in income and output that is larger than the initial reduction in taxes or increase in government spending. The models of the present chapter are intended simply to suggest in very general terms how this expansionary process may work itself out in practice.

chapter 7 13

Foreign Spending

To this point, our analysis has assumed a "closed economy"—an economy considered in isolation from all others. In this chapter, we will relax that assumption and examine an "open economy"—a four-sector economy in which aggregate spending is measured by the spending of the three domestic sectors plus that of the rest-of-the-world sector. In other words, aggregate spending is now defined as the sum of personal consumption expenditures, gross private domestic investment, government purchases, and net exports of goods and services, the last of which is a measure of the amount of aggregate spending for the domestic economy's output that arises from transactions with the rest-of-the-world.

The purpose of this chapter is to describe how this foreign spending for the domestic economy's output affects the level of income and output in the domestic economy. We will confine our analysis to fitting the simple mechanics of exports and imports into an equilibrium model. Just as we omitted detailed analysis of the determinants of government spending and taxing in Chapter 6, here we will omit a detailed analysis of the determinants of an economy's imports and exports. We will adopt a crude theory for the determination of each and then concentrate on how the economy's exports and imports affect the level of income and output and how changes in them cause changes in income and output.

ACCOUNTING IDENTITIES

In Chapter 2 we saw that it is an economy's net rather than gross exports that measure the amount of its final product secured by the rest-of-the-world and the amount of its GNP accounted for by the foreign sector. Gross imports must be subtracted from gross exports in measuring GNP, since gross imports are already included in the measurement of the amount of final product secured by the domestic sectors of the economy. Since consumer, business,

and government expenditures are for both domestic and foreign goods and services, $C + I + G$ no longer measures only the amount of domestically produced final goods and services secured by these three sectors but this amount plus an amount equal to gross imports.[1] If somehow we could separate C into the amounts spent for domestic and foreign output and do the same for I and G, then gross rather than net exports could be added to the resulting total for consumption, investment, and government spending on domestic output to yield aggregate expenditures by all four sectors of the economy for domestic output. Such a complete separation is, of course, impossible in practice. When a consumer purchases a ticket to see a foreign film or purchases a domestically produced automobile that consists in part of raw materials and parts that were produced abroad, he is purchasing simultaneously both domestic and foreign output combined in the same good or service. Similar cases also abound in domestic investment expenditures and in government purchases of goods and services.

In practice, we are therefore forced to measure aggregate expenditures of the three domestic sectors by their expenditures on final product, regardless of whether it was produced domestically or abroad. Such expenditures for any time period will include gross imports of goods and services for that time period. Because this overstates the amount of domestically produced output secured by the domestic sectors, gross imports are subtracted in a lump sum to yield, in effect, the amount of domestically produced output secured by the domestic sectors. Expenditures by the rest-of-the-world for the country's gross exports are then added to the figure obtained by subtracting the country's gross imports from the sum of expenditures by the domestic sectors; the resulting figure is a sum that accurately represents aggregate expenditures, domestic and foreign, for the economy's final output. If I is measured in net terms, we have the following for the net national product or Y identity on the expenditure's side:[2]

$$Y = C + I + G + (X - M)$$

As always, corresponding to the flow of expenditures on the right is an equal flow of income on the left, which is split up as before into C, R_{pf}, S, and T. To show the breakdown on both the expenditures and income sides, we now have

$$C + R_{pf} + S + T = Y = C + I + G + (X - M)$$

Since R_{pf}, personal transfer payments to foreigners, is negligible in terms of net national product, we may simplify the identity by assuming such transfer payments to be zero. This gives us a breakdown of the income flow into $C + S + T$, and the identity becomes

$$C + S + T = Y = C + I + G + (X - M)$$

Although in so doing we no longer have a net national product identity, we may rewrite the preceding identity as

$$C + S + T + M = C + I + G + X$$

The sum of expenditures on the right exceeds net national product by the amount of gross

[1] As in the two- and three-sector economies, I is understood to be net private domestic investment. In the Department of Commerce accounting framework, net private investment is the sum of net private *domestic* investment and net *foreign* investment, and the latter, in turn, is a part of the net export balance as is explained in the Appendix. For present purposes, it is not necessary to show net foreign investment explicitly, i.e., to separate it from net exports of which it is a part. Therefore, we can manage without affixing a subscript to I as would be required if a second investment term, net foreign investment, were shown explicitly.

[2] If we could separate from each domestic sector's expenditures for final product the amount made up of foreign (f) goods and services, symbolically we would have in the four-sector economy

$$Y = (C - C_f) + (I - I_f) + (G - G_f) + X$$

in which X equals gross exports and $C_f + I_f + G_f$ equals gross imports, or M. Since this separation cannot be made in practice, we subtract gross imports in a lump sum from $C + I + G$ and add gross exports:

$$Y = C + I + G - M + X$$

This may then be rewritten in the standard form of the identity that follows in the text above.

imports or M, as does the sum of income on the left. The expression is still an identity, however, and if C is dropped from both sides, we have an identity that is useful in the explanation of the determination of the equilibrium level of income in a four-sector economy.

EQUILIBRIUM LEVEL OF INCOME AND OUTPUT

Other things being equal, the domestic economy's income and output will rise from one period to the next as its gross exports rise or as its gross imports fall, since both these changes enlarge its net exports. It follows that its income and output will fall from one period to the next as its gross exports fall or as its gross imports rise, since both these changes decrease its net exports. It follows in turn that the effect of imports and exports on the equilibrium level of the domestic economy's income and output must be found in the factors that determine the economy's imports and exports.

In a very general way, the volume of any economy's gross exports depends on the prices of goods in the domestic economy relative to the prices of the same or substitute goods in other economies, on tariff and trade policies existing between the domestic and other economies, on the level of income in other countries, on the level of the domestic economy's imports, and on various other less significant factors. Some of the more important factors influencing an economy's exports are not directly related to conditions within that economy. Although this is a great simplification, we will assume that gross exports of the domestic economy are wholly determined by external factors.[3] In other words, we may take gross exports as an autonomous variable, the value of which for any time period is wholly determined by forces outside the domestic economy.

The volume of imports is determined by a similar list of factors. However, many of these factors are much more closely related to conditions within the economy than are those that determine exports. If, in the case of imports, we assume an unchanging system of international price differences, unchanging tariff, trade, and exchange restrictions, and all other factors unchanging except the level of domestic income, we may concentrate on one of the most important factors affecting gross imports—namely, the level of income within the domestic economy.

The Import Function

Other things being equal, as the level of income rises, we expect an induced rise in spending. We may also expect that some portion of the rise in spending will be for imported goods and services. As a rough approximation, we will assume that there is a linear relationship

[3] It should be noted that, with international transactions now included, we must make certain additional assumptions in order to retain the preceding chapters' simplification of a perfectly inelastic aggregate demand curve. Under a system of flexible foreign exchange rates like that now in use, a rise or fall in the domestic economy's price level relative to the price levels of other economies would have to be accompanied by a depreciation or appreciation, respectively, of the domestic economy's currency, other things being equal, or the rise or fall in the domestic economy's price level would itself decrease or increase the gross exports of the domestic economy. This would make total spending on domestic goods vary inversely with the domestic price level or the aggregate demand curve would tilt away from the vertical to slope downward to the right. Different assumptions would be required in the case of a system of fixed exchange rates. In Chapter 18 we will drop the simplification of a perfectly inelastic aggregate demand curve and trace the way that the amount of goods demanded will vary with the domestic price level due to the effect of a change in that price level on the domestic economy's net export balance.

between income and imports, which gives us the import function:

$$M = M_a + mY$$

in which M_a represents autonomous expenditures for imports and m is the *marginal propensity to import,* abbreviated as MPM. As in the $C_a + cY$ function, M_a is the amount of expenditures for imports at a theoretical zero level of income, or the amount of import expenditures that are independent of the level of income, and m is simply the fraction of any change in income that will be devoted to expenditure on imports, or $m = \Delta M/\Delta Y$.

The upward-sloping line shown in Figure 7-1 depicts the import-income relationship described by the import function given above. Since exports are assumed to be externally determined, they are designated in the figure by a line parallel to the income axis. The level of the export line depends on the whole complex of external conditions. For the two functions as illustrated, we see that at all income levels below 0*b* the economy has a net export balance and that at all income levels above 0*b* it has a net import (or negative net export) balance. Clearly, any change in the determinants of gross exports that shifts the export function upward will increase the net export balance or decrease the net import balance at each level of income. Similarly, any change that shifts the import function downward (decreases M_a) or reduces the slope of the import function (decreases m) will have the same effect.

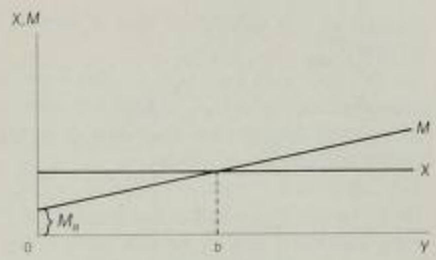


FIGURE 7-1
Import and export functions

The Equilibrium Level of Income—Equations

In the four-sector economy, the equilibrium level of income is that at which aggregate spending—the sum of consumption, investment, government, and net foreign expenditures—is equal to income. This gives us the following general equation for the equilibrium level of income:

$$Y = C + I + G + (X - M)$$

It can also be said that since $C + S + T = C + I + G + (X - M)$, the equilibrium level of income is that at which $S + T = I + G + (X - M)$, or, when it is rearranged,

$$S + T + M = I + G + X$$

In this formulation, $S + T + M$ represents the portion of the economy's gross income flow that is diverted from consumption expenditure on domestically produced output. If the amount of compensating expenditures on domestically produced output, $I + G + X$, just equals these diversions or "leakages" from the income stream, then aggregate spending must equal aggregate income. The particular level of income at which this equality occurs is the equilibrium level.

If we now recognize that the economy's gross imports depend to some extent on the level of its income and assume that its gross exports are externally determined, the equilibrium equation is modified as follows:

$$Y = C + I + G + X - (M_a + mY)$$

We could also say that the equilibrium level of income is that at which

$$S + T + (M_a + mY) = I + G + X$$

Finally, assuming that I, T, and G are all autonomous, and that $C = C_a + c(Y - T)$, the

equation for the equilibrium level of income becomes[4]

$$Y = C_a + c(Y - T) + I + G + X - (M_a + mY)$$

Since $S = -C_a + s(Y - T)$ and $-C_a = S_a$, we may also describe the equilibrium level of income as that level at which

$$S_a + s(Y - T) + T + (M_a + mY) = I + G + X$$

For a numerical example of the determination of the equilibrium level of income, assume the following values for the spending flows:

$$G = 26$$
$$I = 20$$
$$X = 17$$
$$T = 25$$
$$M = M_a + mY = 2 + \tfrac{1}{10}Y$$
$$C = C_a + c(Y - T) = 25 + \tfrac{8}{10}(Y - 25)$$

Substituting in the equation for the equilibrium level of income, we have

$$Y = C_a + c(Y - T) + I + G + X - (M_a + mY)$$
$$Y = 25 + \tfrac{8}{10}(Y - 25) + 20 + 26 + 17 - (2 + \tfrac{1}{10}Y)$$
$$Y = 220$$

or, in its alternative expression:

$$S_a + s(Y - T) + T + (M_a + mY) = I + G + X$$
$$-25 + \tfrac{2}{10}(Y - 25) + 25 + (2 + \tfrac{1}{10}Y) = 20 + 26 + 17$$
$$\tfrac{3}{10}Y = 63 + 3$$
$$Y = 220$$

[4]Apart from the addition of imports and exports, this equation is the same as that developed for the first fiscal model of Chapter 6. If the third fiscal model, which included government transfer payments and the tax function, were expanded to allow for exports and the import function, we would have

$$Y = C_a + c(Y - T_a - tY + R) + I + G + X - (M_a + mY)$$

Alternatively, the equilibrium level of income would be that at which

$$S_a + s(Y - T_a - tY + R) + (T_a + tY - R) + (M_a + mY) = I + G + X$$

The analysis in this text will be limited throughout to the simpler model in which G and T are both autonomous and R is assumed to be zero.

The Equilibrium Level of Income—Graphs

Figure 7-2 illustrates the determination of the equilibrium level of income. Apart from the addition of net exports, the figure is identical in construction to Figure 6-1. The consumption function is plotted from the equation $C = 25 + \tfrac{8}{10}(Y - 25)$. (The vertical intercept is therefore 5.) Autonomous investment of 20 is then superimposed on C to produce the $C + I$ function, and autonomous government purchases of 26 are superimposed on $C + I$ to yield the $C + I + G$ function. Finally, superimposed on the $C + I + G$ function is the last component of aggregate spending, net exports or $X - M$, and this gives the complete aggregate spending function $C + I + G + (X - M)$. Because gross exports are assumed to be independent of the level of domestic income and because gross imports are partially dependent on the level of domestic income, the excess of exports over imports of 15 (i.e., $17 - 2$) at the theoretical zero level of income gradually diminishes to zero at income of 150 [at which $X = 17$, and $M = 2 + \tfrac{1}{10}(150) = 17$]. Therefore, up to Y of 150, foreign trade results in a net addition to aggregate spending—that is, $C + I + G + (X - M) > C + I + G$. Above this level, however, imports exceed exports, and this excess grows larger at successively higher levels of income. Therefore, at income levels above 150, foreign trade results in a net reduction in aggregate spending—that is, $C + I + G > C + I + G + (X - M)$.

Because one component of aggregate spending, net exports, becomes negative at higher levels of income in our example, the aggregate spending function, which includes net foreign spending, lies below the spending function, which is composed of the spending of the three domestic sectors alone, at these higher levels of income.

The equilibrium income level is that level at which the aggregate spending function intersects the 45° guideline. In this case, when $Y = 220$, $C = 181$, $I = 20$, $G = 26$, and $X - M$

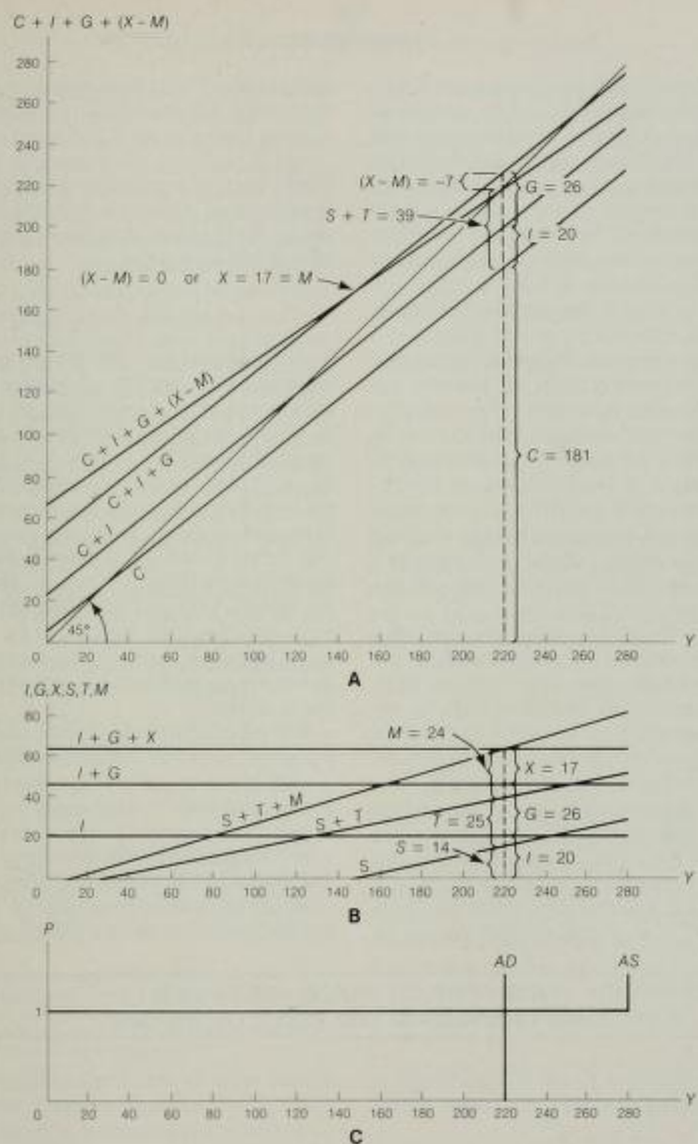


FIGURE 7-2
The equilibrium level of income, including imports and exports

$= -7$. When $Y = 220$, the four components of aggregate spending add up to 220, an amount equal to the income flow; therefore, only 220 can be the equilibrium income level. Following the argument of previous analyses, Part A of Figure 7-2 shows that $C + I + G + (X - M) > Y$ at any level of Y below 220; this excess of spending over income leads to a rise in income and output. At any level of Y above 220, $C + I + G + (X - M) < Y$; this deficiency of spending leads to a fall in income and output.

Given the values we have assumed for the various expenditure flows in our example, the inclusion of foreign trade means a lower equilibrium level than that which would be found in its absence. If $C + I + G$ were to remain the same in the absence of the availability of foreign-produced goods, the equilibrium level would occur at Y of 255, the income level at which the $C + I + G$ function cuts the 45° guideline.[5] The actual equilibrium in our example with foreign trade included is 220.

Part B of Figure 7-2 shows the alternative graphic approach to the determination of the equilibrium income level. Apart from the addition of gross imports and gross exports, the figure is identical in construction to Part B of Figure 6-1. Imports now appear as a third diversion from the amount of expenditure by consumers on domestically produced output; total diversions now equal $S + T + M$. Exports now appear as a third kind of expenditure compensating for the diversions of income into $S + T + M$; total c[illegible]
equal $I +$ [illegible]

foreign trade, the equilibrium level of income is accordingly that at which $S + T + M = I + G + X$. As shown in Part B, this is at $Y = 220$.

Part B of Figure 7-2 enables us to clearly identify each of these diversions from and the compensating injections into the spending stream. Putting them all together, at the income level of 220 we find that

$$S + T + M = I + G + X$$
$$14 + 25 + 24 = 20 + 26 + 17$$

It will be recalled from the analysis of the two-sector economy that if $I > S$, as here $20 > 14$, there would be a disequilibrium, and the level of income would rise. Similarly, in the three-sector economy, if $I + G > S + T$, as here $46 > 39$, there would be a disequilibrium and the level of income would rise. However, in this four-sector economy, the income-expansionary effect of $I + G > S + T$ is offset by the income-contractionary effect of $M > X$, or $24 > 17$. At the income level of 220, these forces offset each other precisely, the sum of the compensating injections matching the sum of income diversions, so income tends neither to expand nor to contract.

With the equilibrium level of income and output shown to be 220 in Parts A and B, the aggregate demand curve is positioned in Part C at that same level. The aggregate supply curve in Part C is the same curve (perfectly elastic to Y of 280) assumed in the preceding chapters. The intersection of the AD and AS curves indi-

mestic income as additional production is turned out to meet increased foreign spending. The MPC indicates that this initial rise in income will induce an increase in consumption expenditures, but the MPM tells us that some of the additional consumption expenditures will be for imported goods. Therefore, at the second stage of the expansion process, domestic income rises, not by the full amount of induced consumption expenditures, but by this amount less the rise in induced consumption expenditures for imported goods. The restricted increase in income at the second stage leads to a smaller third-stage increase in domestic income than would otherwise be the case, for, again, part of the increased expenditures at the third stage is for imported goods. Thus, for any given increase in autonomous spending, the size of the multiplier is reduced when there is a positive marginal propensity to import.

To trace the effect of the marginal propensity to import on the multiplier in terms of our earlier equations, we may begin with the equation developed in this chapter for the equilibrium level of income in the four-sector economy.

$$Y = C_a + c(Y - T) + I + G + X - (M_a + mY)$$

This may be rewritten as

$$Y = \frac{1}{1 - c + m}(C_a - cT + I + G + X - M_a)$$

where $1/(1 - c + m)$ is the foreign trade multiplier for a system in which consumption expenditures and import expenditures are linear functions of the level of domestic income.[6] As in other equations of the same type in preceding chapters, a change in any of the values within parentheses will result in a change in income equal to the change in that value times the multiplier. Let us suppose that there is a change in exports, ΔX.

$$Y + \Delta Y = \frac{1}{1 - c + m}(C_a - cT + I + G + X - M_a) + \frac{1}{1 - c + m}\Delta X$$

Subtracting Y from both sides leaves

$$\Delta Y = \frac{1}{1 - c + m}\Delta X$$

or

$$\frac{\Delta Y}{\Delta X} = \frac{1}{1 - c + m}$$

The same would be true for a change in C_a, I, or G and, with opposite sign, for a change in M_a.[7] Adopting the same values for the MPC and the MPM used earlier in the chapter and assuming ΔX of 18, we have

$$\Delta Y = \frac{1}{1 - \frac{8}{10} + \frac{1}{10}}(18) = 3.33(18) = 60$$

or

$$\frac{\Delta Y}{\Delta X} = \frac{60}{18} = \frac{1}{1 - \frac{8}{10} + \frac{1}{10}} = 3.33$$

If there were no marginal propensity to import or if this propensity were zero, the multiplier would be the ordinary one, $1/(1 - c)$, or, in this example, 5. The rise of 18 in exports would raise the level of income by 90 instead of by 60. Whatever the value of the MPM, as long as it is positive it reduces the size of the effective multiplier. The MPM fits into the determination of the multiplier in the same way as the MPS.[8]

[6] A more complex foreign trade multiplier would emerge from the equation that recognized taxes to be a function of the income level. Adding imports and exports to the equation developed for the third fiscal model of Chapter 6, we have the equation of footnote 4, p. 102.

$$Y = C_a + c(Y - T_a - tY + R) + I + G + X - (M_a + mY)$$

This may be rewritten as

$$Y = \frac{1}{1 - c + ct + m}(C_a - cT_a + cR + I + G + X - M_a)$$

in which the multiplier $1/(1 - c + ct + m)$ is that for the model in which consumption, imports, and taxes are all linear functions of the level of domestic income.

[7] However, for a change in T in the present model, the multiplier becomes

$$\frac{\Delta Y}{\Delta T} = \frac{-c}{1 - c + m}$$

See p. 87.

[8] In these terms, $1 - c$ is the MPS, or s. The greater $1 - c$, the greater is s, the greater is the "leakage" into saving from any change in income, and therefore the smaller is the expansion in income for any increase in

Another way of illustrating the effect of the MPM on the multiplier is to express the multiplier as $1/[1 - (c - m)]$, in which, as before, c is the marginal propensity to purchase both domestically produced goods and foreign-produced goods and m is the marginal propensity to purchase foreign-produced goods. The expression $c - m$ accordingly represents the marginal propensity to purchase domestically produced goods and is relevant to changes in the domestic level of income. If m were equal to c, the multiplier would be 1, for in this case any rise in autonomous spending would raise income only by the amount of that rise in autonomous spending. The full amount of the increase in income received at the first stage would be diverted to the purchase of foreign produced goods; there would be no induced increase in the purchase of domestically produced goods. As long as m is less than c, which it usually is, the multiplier will exceed 1. And finally if m were zero, c would become identical to the marginal propensity to purchase domestically produced goods, or the multiplier $1/[1 - (c - m)]$ would in effect be the ordinary multiplier $1/(1 - c)$.

Figure 7-3 depicts the method of determining the change in the equilibrium level of income that results from a change in autonomous spending.[9] In Part A, for the income range shown, the solid-line schedules are the same as those in Figure 7-2. The aggregate spending function, $C + I + G + (X - M)$, intersects the 45° line at the income level of 220 as in Figure 7-2. Now, suppose that there is an increase of 18 in exports.[10] The resultant aggregate spending function, now labeled $C + I + G + (X + \Delta X - M)$, lies 18 above the previous aggregate spending function at each level of income and intersects the 45° line at income of 280, the new equilibrium level. With a rise of 18 in autonomous spending and a consequent rise of 60 in income and output, the multiplier is $\Delta Y/\Delta X = 60/18 = 3.33$, as was determined earlier in terms of the equation for the equilibrium level of income.

With the equilibrium at 280, the perfectly inelastic *AD* curve of Part C is positioned at Y of 280, which is the full employment level of output as indicated by the perfect inelasticity of the *AS* curve at this level of output. As noted for the full employment case in the preceding chapter, the coincidence of the *AS* and *AD* curves at all possible price levels (although they are shown slightly apart in the diagram for identification) means that the price level is indeterminate. However, the equilibrium level of income and output, which is our present concern, is established at Y of 280 as shown.

Before exports rose by 18, foreign trade exerted a net contractionary effect on the income level, for at the equilibrium level of 220, $X - M = 17 - 24 = -7$. The aggregate spending function lay below the spending function for the three domestic sectors alone. After the rise of 18 in exports and the establishment of the new equilibrium of 280, foreign trade exerts a net expansionary effect. At this new income level, $X + \Delta X - M = 17 + 18 - 30 = 5$. The new aggregate spending function now lies above the spending function for the three domestic sectors. Notice, however, that the expansionary effect following the rise in exports is checked to some extent by the induced expenditures for foreign-produced output. The ordinary multiplier of 5, indicated by the MPC of 8/10, would produce a rise of 90 in income with ΔX of 18.

---

autonomous spending. In turn, $s + m$ is the sum of the MPS and the MPM. The greater this sum, the greater is the "leakage" into saving and imports from any change in income, and therefore the smaller is the expansion in income for any increase in autonomous spending. Finally, for the multiplier including the MPT, ct is added to $s + m$ to produce $s + ct + m$. This, then, is the sum of MPS, MPC × MPT, and MPM. The greater this sum, the greater is the "leakage" of any change in income into saving, imports, and taxes, and the smaller is the expansion in income for any increase in autonomous demand.

[9] In order to show better detail over the relevant range of income, note that this figure shows each function only over this narrower range of income.

[10] This unrealistically large increase is chosen to avoid "congestion" in the figure.

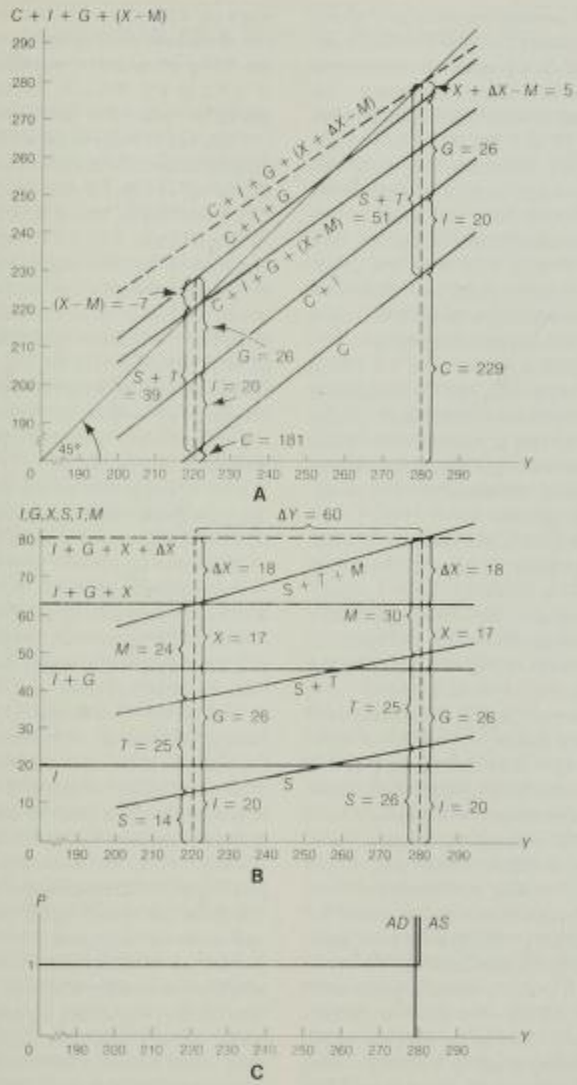


FIGURE 7-3

Effect of a change in exports on the equilibrium level of income

Of course, on the present assumption that the aggregate supply curve becomes perfectly inelastic at Y of 280, this greater rise in income would turn out to be purely inflationary. The aggregate demand curve in Part C would shift to a position at Y of 310; the 30 excess of aggregate demand over the upper limit of 280 for aggregate supply would have to be absorbed by a rise in the price level. However, if we change our assumption and put the full employment output at Y of 310, then the larger increase in real income of 90 that otherwise could be obtained would be prevented by the MPM of 1/10, which reduces the effective multiplier to 3.33 and thus holds to 60 the rise in income from ΔX of 18. This income-restraining effect of the MPM is depicted graphically as the narrowing of the spread at successively higher levels of income between the spending function of the three domestic sectors and the new aggregate spending function.

Part B of Figure 7-3 shows these two equilibrium positions expressed in terms of the equality between $S + T + M$ and $I + G + X$. Again, apart from the narrower income range shown here, the solid-line functions are the same as those shown in Part B of Figure 7-2. There the initial equilibrium, $S + T + M = I + G + X$, occurred at an income of 220. The increase of 18 in exports shifts the entire $I + G + X$ function, labeled $I + G + X + \Delta X$, upward by 18. The increase in exports, ΔX, sets the multiplier in operation and raises income to the new equilibrium level of 280. Since in the present model I, G, and T are assumed to remain unchanged despite changes in income, the new equilibrium must occur at that level of income where $\Delta X = \Delta S + \Delta M$. In other words, given the injection of ΔX, Y must rise to the level at which the diversions, $\Delta S + \Delta M$, from the income stream match the injection of ΔX into the income stream. Since the MPS = 2/10 and the MPM = 1/10, it is only when income rises by 60 that the sum of ΔS, which is $\frac{2}{10}(60)$, or 12, and ΔM, which is $\frac{1}{10}(60)$, or 6, will match ΔX. The rise in income must be 3.33 times the rise in exports, as we can see by expressing the multiplier in terms of the MPS, or s, and the MPM, or m: $\Delta Y/\Delta X = 1/(s + m)$. Thus, in the present example, $60/18 = 1/(2/10 + 1/10) = 3.33$.

If the rise in income generated by the rise in exports did not induce expenditures for foreign goods as well as for domestic goods, the net effect on income and output (assuming again for the moment that the full employment level of output is 310) would be an increase of 90, or the increase of 18 in exports times the ordinary multiplier of 5 indicated by the MPS of 2/10. The income-restraining effect of the MPM is shown in Figure 7-3 as the widening spread at higher income levels between the schedule of "leakages" labeled $S + T$ and the schedule of "leakages" including imports labeled $S + T + M$.

All this suggests that the greater the MPM, the greater the reduction in the effective multiplier, and therefore the smaller the expansion of income that follows any specific increase in autonomous spending. In practice, the MPM shows considerable variation among countries. It tends to be higher in countries that are heavily engaged in foreign trade, such as England, than in countries that are not, such as the United States. This suggests further that, with allowance for any difference in the MPC, any increase in autonomous spending will have a smaller multiplier effect in a country such as England than in a country such as the United States. However, this conclusion is based on our assumption that a country's exports are entirely determined by external factors. If we drop this assumption and recognize that internal as well as external factors affect a country's exports, the probability that a country like England will have a smaller effective multiplier than a country like the United States does not necessarily follow. To see why this is so, we will consider in the following section one of the ways in which a nation's exports are tied to domestic factors.

EXPORTS AS A FUNCTION OF IMPORTS

In any country operating below its full employment level of output, an increase in consumption, investment, or government spending will raise the level of output and income as long as there are no offsetting decreases in its net export balance or increases in its net import balance. With a positive marginal propensity to import, however, the rise in output and income leads to a rise in imports and, in the first instance, to a decrease in the net export balance or an increase in the net import balance. Such a rise in, say, U.S. gross imports is felt by all other countries combined as an equal increase in their gross exports. Since the rest-of-the-world experiences an increase in gross exports with no simultaneous increase in gross imports (U.S. exports did not increase in the first instance), one or more countries in the rest-of-the-world must show an increase in net exports or a decrease in net imports. This means that one or more countries will, other things being equal, show a rise in their domestic income levels. But these countries also have a positive marginal propensity to import, and some portion of the increase in their incomes will be diverted to the purchase of imported goods and services. There is reason to expect that the U.S. economy will secure some share of the increase in purchases made abroad by these foreign countries, and this will appear as a rise in U.S. gross exports. In this complicated fashion, changes in U.S. gross *exports* are indirectly influenced by changes in the income level in the United States.

If we recognize that changes in the level of income in the United States are a factor determining changes in U.S. exports, the earlier assumption that exports are determined entirely by external factors appears to be invalid. This assumption is nonetheless a valid approximation for small economies, in which even a large percentage change in income will be relatively small in absolute terms. Even with a high marginal propensity to import, the effect of such an increase in income on the country's gross imports will not be large enough to affect significantly the aggregate exports of other countries and thus the level of income in those countries. This being the case, the small country cannot expect an appreciable feedback in the form of increased exports to result from the increase in its imports. We must thus distinguish between countries that are large enough to influence perceptibly the income levels of other countries through changes in their imports and countries that are too small to do this. Size in this context is not measured merely by the real income level of the country; it is a compound of income level and marginal propensity to import. In terms of our earlier comparison, although the U.S. marginal propensity to import is smaller than England's, the relatively high income level of the U.S. economy means that a moderate percentage change in U.S. income may have a noticeable effect on the exports of the rest-of-the-world. (A 6-percent rise in GNP runs to about \$100 billion at annual rate; an MPM of 1/20 of GNP then means added imports of about \$5 billion at annual rate.) On the other hand, although the income level of England is far smaller than that of the United States, the higher marginal propensity of the English economy to import means that even a moderate percentage change in its income level can have a noticeable impact on exports of the rest-of-the-world. (Based on the recent foreign exchange rate of £1 = \$1.70, a 6-percent increase in GNP means roughly \$15 billion at an annual rate; an MPM of 1/5 of GNP means added imports of roughly \$3 billion at an annual rate.)

When there is a vigorous business expansion

under way in a large country like the United States, the rising level of income induces an increased flow of imports into the United States. The most recent illustration of this is what happened in the United States during the recovery that followed the severe recession of 1973–75. Imports increased from $127.6 billion in 1975 to $155.1 billion in 1976 or by 22 percent. The increased flow of imports is felt by some countries in the rest-of-the-world as a rise in their incomes. This induces a rise in imports in these countries, and, to the extent that the United States shares in these expanding foreign markets, U.S. exports increase. This further raises the income level in the United States and further enlarges U.S. imports from the rest-of-the-world. These repercussions continue to interact on the income levels of both the United States and countries in the rest-of-the-world. However, since the marginal propensities to import both here and abroad are less than 1, the amount of spending at each succeeding stage of this expansion process will decrease until income in the United States and in other countries tends to stabilize at new levels. In practice, of course, income changes of this sort are continually occurring as the result of changes in domestic spending in various countries. Therefore, before one series of repercussions can work itself out, a new series is initiated by further changes in domestic spending in one or more important countries.

This crude description suggests why the income levels of different countries are interdependent. A rising level of income in a large country like the United States tends to raise income levels of some other countries. By the same process, a falling level of income here is felt by other countries as shrinkage in their exports and declining levels of income. As nations become more and more closely tied together through foreign trade, we encounter what economists have called the "international propagation of business cycles." Depression in one or more large countries tends to trigger depressions in other countries; prosperity in one or more large countries tends to promote prosperity in other countries.[11]

A CONCLUDING NOTE

As we mentioned at the outset, the purpose of this chapter has been to describe the way in which an economy's net export or import balance enters into the determination of that economy's equilibrium income level and the way in which changes in that balance can cause the economy's income level to change. The basic model was developed on the assumption that exports are externally determined and imports internally determined, and specific attention was paid to only one of the many internal factors that influence imports—namely, the level of income. Although we did not develop a formal model, we did note the way in which the exports of a large nation may depend indirectly on its level of income.

Many other factors also influence a nation's net export or import balance and thereby influence its income level. In a more thorough analysis, for example, we would recognize that differences in the relative prices of goods in different countries are the reason for international trade in such goods. If an economy finds its price level rising relative to the price levels of other economies, it may expect its imports to rise, even if its marginal propensity to import and its real income remained unchanged. (In

[11]A formal algebraic statement of what has been loosely stated here will be found in books in international economics. See, for example, M. E. Kreinen, *International Economics, A Policy Approach*, Harcourt Brace Jovanovich, 2nd ed., 1975, Appendix I, pp. 409–13, and C. P. Kindleberger, *International Economics*, 5th ed., Irwin, 1973, Appendix G.

terms of the import function, this would appear as an increase in M_a). At the same time that its imports rose, its exports would decline as foreign importers (like domestic buyers) shifted purchases to other countries that offered the desired products at lower prices. To show how this fits into a model clearly requires a model in which price level is treated as a variable. This is done in Chapter 18 which brings the foreign sector into an extended model with a variable price level.

A more thorough analysis would also have to consider variations in currency exchange rates, a factor that has become increasingly important in the years since 1973 when the foreign exchange values of many of the world's leading currencies were left to float more or less freely in response to short-run changes in their supply and demand. Over the period from the end of World War II until the turn to floating rates in the early seventies, the Bretton Woods system had provided short-run exchange rate stability as nations supported exchange rates within a prescribed range. In the judgment of many observers, the experience with floating rates in recent years, which includes a period of severe international recession, provides substantial evidence of the workability of such a system. This suggests that as the international monetary system evolves in the years ahead, it will continue to be one that provides much greater flexibility in exchange rates than was provided during the preceding quarter of a century under the Bretton Woods system. What is relevant at this point is that an international monetary system in which exchange rate adjustments become more frequent would make changes in exchange rates a more important influence on the net import and export balances of nations than they have been in the past. The fluctuation of exchange rates therefore becomes another important factor that must be taken into account in any analysis that purports to be at all thorough. An introduction to the way that changes in the foreign exchange rates affect the economy's income and output levels is provided in the extended model in Chapter 18.

Some other factors that are also at work are changes in tariffs, quotas, foreign-exchange controls, and other controls imposed by each economy. The effects on the net import or export balance that might otherwise follow from changes in price levels or in foreign-exchange rates could, at least temporarily, be offset by appropriate manipulation of these devices.

A thorough analysis of these and other such factors may be found in basic texts on international economics. We mention them here simply to emphasize the rigid assumptions on which the simple analysis in this chapter is based. Yet this analysis has given us some insight into the question with which we started: How do an economy's foreign transactions affect the level of income and output within that economy?

part three

The Theory of Consumption and Investment Spending

chapter 8 The Level of Income and Consumption Spending

In developing the simple Keynesian model of income determination in Part 2, we worked with a consumption function in which the MPC was constant, positive, and less than 1 and in which the APC exceeded the MPC at all levels of income. Although it was suggested that such a relationship between consumption and income appeared to be plausible, no real evidence was submitted in its support; the relationship was simply derived from a particular theory of consumer behavior that was described briefly in Chapter 4. The major tenets of this theory are that consumers devote a fraction of any increase in income to saving and that they save a larger fraction of a higher income than of a lower income. This specific theory of consumer spending, apart from some simplifications of a noncritical nature, is essentially the view that Keynes advanced in his *General Theory*.

Keynes's theory has been elaborated by others and has come to be known as the *absolute income hypothesis* or theory. Rival theories have also been developed, the major two being the *relative income* and *permanent income hypotheses* or theories. Each of these three theories of consumer spending involves a quite specific but different theory of consumer behavior—that is, each assigns a different role to income as an influence on consumer spending, or, otherwise expressed, each sees income as a determinant of consumption in a somewhat different way.

The primary purpose of this chapter is to investigate the relationship between consumption and income through an examination of these three major theories of consumer spending. The economists who have constructed these theories have all begun with a theory of individual consumer behavior and then generalized to cover aggregate behavior. In the testing of these theories, cross-sectional data provide empirical evidence on how spending varies at different levels of family income in any one year, and time series data provide empirical evidence on how aggregate spending (or spending by all families combined) varies as aggregate income (or the income of all families combined) changes from one year or one quarter to the next. Our procedure in this chapter will follow a similar line. First, we will discuss the absolute income, relative income, and permanent income theories as theories of individual consumer behavior and relate them to some simple cross-sectional data; then we will discuss the same three theories in their generalized, or aggregative, form and relate the

generalized theories to the aggregate, or time series, data. This procedure divides our task into two parts that may be called microanalysis and macroanalysis.

CONSUMER BEHAVIOR: MICROANALYSIS

At the micro level, the relevant data are those that show how much of the individual family's income, on the average, is devoted to consumption and saving for a sample of families at various income levels. The most recent source of such data is the 1972–73 Consumer Expenditure Interview Survey conducted by the U.S. Department of Labor. However, this survey did not gather data which show consumption by after-tax income class, only consumption by pre-tax income class. For present purposes, only consumption by after-tax income class is meaningful, and for a study which provides this we must go back to an earlier 1961 Department of Labor survey. Table 8-1 summarizes the data that are relevant here. Columns (2) and (3) give, in dollars, the average disposable income and the average consumption expenditures of the sample families in each income class listed in column (1). Individual sample families within each income class are, of course, scattered above and below the indicated averages for disposable income and consumption expenditures. Furthermore, some individual families within any income class may be below the average in terms of income and above the average in terms of consumption and some others just the other way around, but it is the average and not the individual variations to which we must look to discern whatever pattern exists.

Column (4) was derived by dividing the dollar amounts in column (3) by the dollar amounts in column (2). The values in column (4) may be related to our earlier concept of the average

apc = Con. Exp / Dis Inc.

TABLE 8-1
Average disposable income and consumption expenditures by income class, all urban and rural families and single consumers, 1961

(1) INCOME CLASS	(2) DISPOSABLE INCOME	(3) CONSUMPTION EXPENDITURE	(4) apc	(5) mpc
Under $1,000	$ 524	$ 1,291	2.46	
$1,000– $1,999	1,520	1,766	1.16	0.48
$2,000– $2,999	2,512	2,679	1.07	0.92
$3,000– $3,999	3,505	3,572	1.02	0.90
$4,000– $4,999	4,494	4,346	0.97	0.78
$5,000– $5,999	5,485	5,179	0.94	0.84
$6,000– $7,499	6,717	6,053	0.90	0.71
$7,500– $9,999	8,555	7,382	0.86	0.72
$10,000–$14,999	11,761	9,507	0.81	0.66
$15,000 and over	22,144	14,273	0.64	0.46
All income classes	5,594	5,038	0.90	

SOURCE: *Consumer Expenditures and Income, Total United States, 1960–61*, BLS Report No. 237–93, Feb. 1965, Table 1-B, p. 16. Columns (4) and (5) were derived by the author from columns (2) and (3).

mpc = —— Δ Avg. Inc

propensity to consume if we recognize, however, that these values relate to different levels of family income and not to different levels of aggregate income. The values in column (4) show that as we move from lower to higher family incomes, the percentage of income devoted to consumption expenditures, the *apc*, decreases. (To differentiate the family from the aggregate average propensity to consume, *apc* is used for the former and, as before, APC for the latter.) Column (5) was also derived from the dollar amounts in columns (2) and (3). Here the change in average consumption between each pair of income classes is divided by the change in average income between the corresponding pair of income classes. Again recognizing that the values relate to different levels of family income rather than to different levels of aggregate income, we may relate the values in column (5) to our earlier concept of the marginal propensity to consume. These values show that as we move from lower to higher income classes, consumption expenditures increase but by less than the increase in income. With minor variations, other family budget studies made at different times reflect the same properties: The *apc* decreases with higher levels of family income, and the *mpc* (here using *mpc* for the family marginal propensity to consume) is positive but less than 1.

The data given in columns (2) and (3) of Table 8-1 have been plotted in Figure 8-1. The broken vertical lines indicate the limits of the income classes of Table 8-1; the dot between each pair of broken lines indicates the average disposable income (from column (2)) and average consumption expenditure (from column (3)) for the sample of families within each income class so delineated. These dots have been connected by a smooth line that may be called the *family consumption function*.

The familiar 45° line has been inserted in Figure 8-1 as a guide. Since the family consumption function crosses the 45° line at a family income level in the neighborhood of $4,000, we may deduce that in 1961, on the average, a family income level of about this amount represented the break-even income, or the income level at which consumption expenditures were equal to income so that there was neither family saving nor family dissaving. At successively lower family income levels, there was increasing dissaving by the average family. Conversely, at successively higher family income levels, there was increasing saving by the average family. Apart from some curvature of the family consumption function of Figure 8-1, which indicates that the *mpc* was not constant, there is a general similarity between the position and shape of this function and of those in Part Two. They reflect the same properties: decreasing dissaving as we move from very low income to the break-even income and then increasing saving as we move up to higher income.

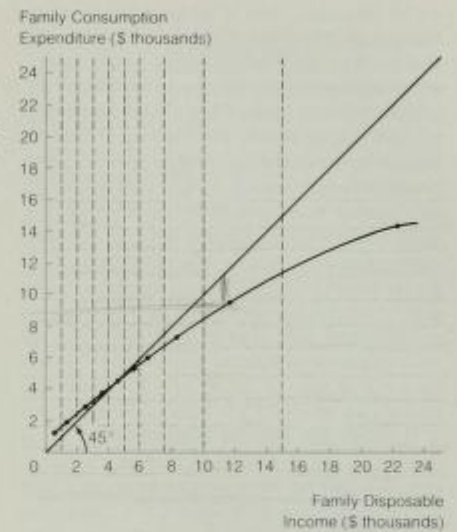


FIGURE 8-1
Family consumption function, 1961

Having demonstrated that, apart from some curvature, the properties of the empirically derived family consumption function are the same as those of the aggregate consumption function

earlier assumed, one is tempted to conclude that the latter is a reasonably accurate description of the way that aggregate consumption expenditures actually do vary as the level of aggregate income varies. This simple conclusion is ruled out, however, when we recall that we are talking about two completely different types of consumption function. One shows for a given time period how families at different income levels divide their respective incomes between consumption and saving. The other shows how all families combined would divide different *alternative* levels of aggregate income between consumption and saving. Whether or not we may apply conclusions drawn from cross-sectional studies of *family* income differences to the case of *aggregate* income differences is a question that can be approached only by turning to the theory of individual consumer behavior and the various theories of consumer spending mentioned above.

The Absolute Income Theory

The basic tenet of the absolute income theory is that the individual consumer determines what fraction of his current income he will devote to consumption on the basis of the *absolute* level of that income. Other things being equal, a rise in his absolute income will lead to a decrease in the fraction of that income devoted to consumption. As we have noted, the first statement of this hypothesis is probably that made by Keynes in the *General Theory*. Its subsequent development is primarily associated with James Tobin and Arthur Smithies.[1]

For a simple illustration, assume that we were able to take all families at three different income levels, specifically $4,000, $8,000, and $12,000, for a given year and to ascertain the average consumption expenditures of all families at each of these income levels. We would expect these average consumption figures when plotted to show up as points like *D*, *E*, and *F* on the curve labeled *C* in Figure 8-2. (Disregard the curve labeled *C*′.) A pattern such as this for the three points is clearly what cross-sectional data from actual budget studies typically show. Suppose next that all families at these income levels, as well as all families at all other income levels, somehow double their income between this initial year and a subsequent year. Assuming no other changes, how do families at each income level, on the average, divide their doubled incomes between consumption and saving? The theory of behavior that underlies the absolute income theory suggests that families would, on the average, divide their now doubled incomes in the same way as did the average family that previously occupied that income position. This would mean that the families at the three income levels mentioned above would move along the family consumption function in Figure 8-2 from *D* to *E*, from *E* to *G*, and from *F* to *H*, respectively. Families at other income levels would move along the function in a similar manner. Because it is the absolute level of the family's income that is held to control the division of income between consumption and saving, this theory predicts a decline in the average propensity to consume of the average family when families move up to a higher income level.[2]

If consumers do, in fact, respond to changes in income in this way, it follows that, of the aggregate income received by all consumers, the fraction spent decreases as the aggregate income increases. This relationship between aggregate income and aggregate consumption

[1] See J. Tobin, "Relative Income, Absolute Income, and Saving," in *Money, Trade, and Economic Growth*, Macmillan, 1951, pp. 135–56, and A. Smithies, "Forecasting Postwar Demand: I," in *Econometrica*, Jan. 1945, pp. 1–14.

[2] Thus, the $4,000 family whose apc was, on the average, 1.2 (or $4,800/$4,000) now becomes an $8,000 family whose apc is, on the average 1.0 (or $8,000/$8,000). The $8,000 family whose apc was, on the average, 1.0 ($8,000/$8,000) now becomes a $16,000 family whose apc is, on the average, 0.8 (or $12,800/$16,000). And the $12,000 family whose apc was, on the average, 0.9 (or $10,800/$12,000) now becomes a $24,000 family whose apc is, on the average, 0.6 (or $14,400/$24,000).

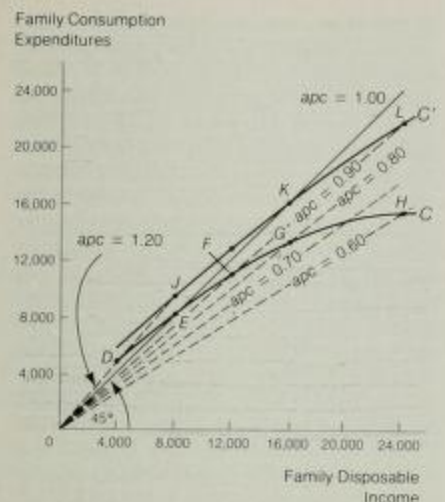


FIGURE 8-2
Family consumption function: absolute and relative income

is one of the properties of the aggregate consumption function we employed in Chapter 4. That function, it will be recalled, was drawn to correspond, in the main, with the theory of consumer behavior stated by Keynes. Since that theory was essentially what we now call the absolute income theory, the relationship noted here, in which the fraction of aggregate income consumed declines as aggregate income rises, is of course the one to be expected.

The absolute income theory seems to be a plausible theory of individual consumer behavior, and the noneconomist would probably accept it today without hesitation. If he were presented with the question before us, he would most likely say, in plain language, that high-income families save a larger fraction of their incomes than low-income families for the obvious reason that they have a relatively large fraction of their incomes left after meeting consumption needs, whereas low-income families find that consumption needs absorb or even exceed their total incomes. To him it would then be equally obvious that as families move up the income scale, they are able to save a larger fraction of their income at each higher income level. In the first years following the appearance of the *General Theory*, economists also generally accepted the absolute income theory as basically correct, but the widespread acceptance enjoyed by this theory was short-lived. Although it still has some supporters, most economists lean toward one of the other theories, such as the relative income theory. Some reasons for this will become apparent as we continue.

The Relative Income Theory

The relative income theory, which is closely associated with the name of James S. Duesenberry,[3] argues that the fraction of a family's income devoted to consumption depends on the level of its income *relative* to the income of neighboring families or other families with which it identifies and not on the *absolute* level of the family's income. Thus, if a family's income rises but its *relative position* on the income scale remains unchanged because the incomes of other families with whom it identifies have risen at the same rate, its division of income between consumption and saving will remain unchanged. The family's absolute income has risen, so its asolute consumption and absolute saving will also rise, but the *fraction* of income devoted to consumption will be the same at the higher level of absolute income as it was at the lower level. To express the same argument in a different way, if a family's income remains unchanged but the incomes of these

[3]See J. S. Duesenberry, *Income, Saving, and the Theory of Consumer Behavior*, Harvard Univ. Press, 1949. See also D. S. Brady and R. Friedman, "Savings and the Income Distribution," in *Studies in Income and Wealth*, Vol. 10, National Bureau of Economic Research, Princeton Univ. Press, 1947, pp. 247–65, and F. Modigliani, "Fluctuations in the Saving-Income Ratio: A Problem in Economic Forecasting," in the same series, Vol. 11, 1949, pp. 371–443.

other families rise, its income position relative to that of the other families has then changed. The relative income theory would argue that the deterioration in the relative position of this family would lead to a rise in the fraction of its income devoted to consumption, despite the fact that there has been no change in its absolute income.

In its focus on relative income, this theory emphasizes the imitative or emulative nature of consumption. A family with any given level of income will typically spend more on consumption if it lives in a community in which that income is a relatively low one than if it lives in a community in which that income is a relatively high one. This tendency to spend more in the one situation arises in part from the pressures on the family to "keep up with the Joneses" and in part from the fact that the family will observe in its everyday living what to it are the superior goods of other families and will be tempted to spend as a result of what Duesenberry calls the "demonstration" effect. Thus, studies have shown that black families with an income of, say, \$10,000 will, on the average, spend less than white families with an equal income, the reason being that the black family with this income will most likely live in a neighborhood in which \$10,000 is a relatively high income, whereas the white family will most likely live in a neighborhood in which this same income is farther down the scale of incomes of families in that neighborhood.

If families behave in the way indicated by the relative income theory, the division of their income between spending and saving will clearly be different from the division suggested by the absolute income theory. This basic difference may be seen by turning back to Figure 8-2. Assume as before that the absolute income of all families doubles between an initial year and a subsequent year. There is then no change in the distribution of income. The top 1 percent of all families gets the same percentage of the aggregate income in both years, the bottom 1 percent gets the same percentage in both years, and similarly for every other percentile. Therefore, despite the doubling of each family's income, its relative position on the income scale is the same as it was before. The relative income theory then argues that there would be no reason, on the basis of the income change alone, to expect a change in the fraction of income consumed by the average family. If there is no change in this ratio, with the doubling of its income each family would not move upward along the function labeled C, as suggested by the absolute income theory, since such a movement indicates a decrease in the ratio of income spent by each family. The relative income theory argues instead that each family moves in such a way as to create the new consumption function labeled C'. That is to say, the \$4,000 family would move not from D to E but from D to J; the \$8,000 family would move not from E to G but from E to K; and the \$12,000 family would move not from F to H but from F to L. If the average family at each income level did respond to this doubling of income in the way indicated by the relative income theory, its *apc* would be the same before and after the doubling of its income. Each family is on the same *apc* curve (depicted in Figure 8-2 by broken straight lines drawn from the origin) that it was on before the rise in its income.[4] According to the relative income theory, each family, in deciding on the fraction of its income to be spent, is uninfluenced by the fact that it is twice as well off in *absolute* terms and is influenced only by the fact that it is no better off at all in *relative* terms. Being no better off in this sense, its decision is to "live" as it did previously, devoting the same fraction of its income to consumption as before.

The relative income theory reaches this conclusion on the assumption that the distribution

[4] Thus, the \$4,000 family whose apc was 1.2 (or \$4,800/\$4,000) is now an \$8,000 family whose apc is still 1.2 (or \$9,600/\$8,000). The \$8,000 family whose apc was 1.0 (or \$8,000/\$8,000) is now a \$16,000 family whose apc is still 1.0 (or \$16,000/\$16,000). And the \$12,000 family whose apc was 0.9 (or \$10,800/\$12,000) is now a \$24,000 family whose apc is still 0.9 (or \$21,600/\$24,000).

of income remains essentially unchanged as the level of aggregate income changes. However, if a redistribution occurs, those families whose incomes rise less rapidly than the average of those with whom they identify will tend to raise their *apc*, while those whose incomes rise more rapidly will tend to lower their *apc*. If increases in income are accompanied by a redistribution toward greater equality, the *apc* for all families (with the exception of those at the top of the income scale) will tend to be reduced. This again follows from the argument that the fraction of family income consumed is determined by the family's relative position on the income scale. The pressure of families at each level of the income scale to "keep up with the Joneses" is reduced as income differentials are reduced through redistribution toward greater equality.[5] Such redistribution brings the income of any one family and those families that it emulates or imitates closer together and thereby tends to reduce this family's *apc* by reducing the amount of its consumption that is imitative in origin.

The distribution of income in the United States has been altered in the direction of greater equality over the long run, but the changes have been so gradual that appreciable change has been apparent only over a period of several decades or more.[6] On the other hand, the growth in real income itself has been quite rapid, per capita GNP having increased by 60 percent from 1950 to 1975. Thus, when changes in distribution are taken into account, we cannot expect the family consumption function to shift from C to C′ in Figure 8-2 in the simple way described earlier. Still, the actual changes in income distribution do not invalidate but only qualify the result that would be expected if the level of average family income increased with no change in income distribution.

The relative income theory, like the absolute income theory discussed earlier and the permanent income theory to be discussed next, is a theory of consumer behavior that relates the consumer's spending to his income on the assumption that there is no change in any of the other factors influencing consumption, of which the distribution of income is only one. Under this assumption, the relative income theory gives us the result described above: that the family consumption function shifts upward in proportion with changes in aggregate income received by all families. This result, therefore, turns out to be inconsistent with the result indicated by the absolute income theory. If relative income controls the fraction of income spent by the family, an equal percentage change in the incomes of all families will produce no change in the fraction of that enlarged aggregate income that is devoted to consumption. On the other hand, if absolute income controls it, the fraction of aggregate income consumed declines as aggregate income increases. There will be a further comparison of the results reached by the absolute and relative income theories in the time series analysis later in this chapter.

The Permanent Income Theory

Both the absolute and relative income theories focus on the individual family's "current" income as the income concept relevant to its spending. What is the meaning of current income in this context? Is the family's current income its measured income for a week, a month, a year, two years? Does the family adjust its consumption upward or downward from one week to the next in the face of a rise or fall in its measured income from one week to the next? Or does it do this on a monthly basis or a yearly basis? It seems fair to say that a family's spending in any one week is not closely related to its measured income during that specific week; we would not expect its consumption to be drastically affected in any one week even if its measured income for that one week alone should be zero. The same is true to a lesser

[5] The influence of income distribution on consumption is examined further in Chapter 9.

[6] For some evidence, see footnote 14 on p. 148.

Perm vs Transitory

degree if we take a time span of one month. Where is the line to be drawn? Most studies that take current income as the appropriate income concept take the current year, or sometimes the current and the preceding year, as the relevant time span.

A quite different approach to the role of income in the theory of consumer spending has been developed by Milton Friedman. His point of departure is the rejection of the usual concept of "current" income and its replacement with what is called "permanent" income.[7] A family's permanent income in any one year is in no sense indicated by its current income for that year but is determined by the expected or anticipated income to be received over a long period of time, stretching out over a number of future years. In Friedman's words, permanent income "is to be interpreted as the mean income regarded as permanent by the consumer unit in question, which in turn depends on its horizon and farsightedness."[8] The time span that is relevant to permanent income is the minimum period of time over which income influences must be maintained in order to make the receiver of that income regard them as permanent.

Given this meaning of permanent income, a family's measured or observed income in any particular year may be larger or smaller than its permanent income. Friedman divides the family's measured income in the year into permanent and transitory components, so that its measured income is larger or smaller than its permanent income, depending on the sum of positive and negative transitory income components. For example, if a family wage earner receives an unexpected special bonus at work in one year and has no reason to expect the same bonus in following years, this income element is regarded as positive transitory income, and it raises his measured income above his permanent income. On the other hand, if he suffers an unexpected loss of income due, say, to a plant shutdown as a result of fire, this income element is regarded as negative transitory income, and it reduces his measured income below his permanent income. These unforeseen additions to and subtractions from a family's income are expected to cancel out over the longer period relevant to permanent income, but they are present in any shorter period.

In the same way, Friedman divides measured consumption into permanent and transitory components.[9] A good purchased because of an attractive sale price or a normal purchase deferred due to unavailability of the good would be examples of positive and negative transitory consumption. As with measured income, a family's measured consumption in any particular period may be larger or smaller than its permanent consumption.

With these definitions at hand, we may turn to Friedman's basic argument that permanent consumption depends on permanent income. Specifically, the relationship he proposes is that permanent consumption is a constant proportion of permanent income in which the proportion depends only on the interest rate, on the ratio of "nonhuman" wealth to total (human plus nonhuman) wealth, and on tastes. Tastes are affected by factors such as age and family composition. The permanent consumption of

[7]See M. Friedman, *A Theory of the Consumption Function*, Princeton Univ. Press, 1957. A very similar theory, sometimes called the "life-cycle" hypothesis, was developed more or less independently by F. Modigliani and others. See A. Ando and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," in *American Economic Review*, March 1963, pp. 55–84, and F. Modigliani and R. E. Brumberg, "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. K. Kurihara, ed., *Post-Keynesian Economics*, Rutgers Univ. Press, 1954, pp. 388–436. Although the two are very similar in principle, the Friedman version has gained wider attention, and discussion here will center on that version. The fact that Friedman uses the word *theory* and Modigliani uses the word *hypothesis* does not indicate a difference in the validity or generality of the two. It merely reflects a choice of words.

[8]Friedman, *op. cit.*, p. 93.

[9]It should also be noted that Friedman defines consumption as spending on services and nondurable goods plus the depreciation of consumer durable goods. A net addition to the family's stock of durable goods is treated as saving.

different families with the same permanent income will thus vary with their tastes and other specified characteristics. However, if there is no reason to expect these characteristics to vary with the level of income, it may be assumed that the average ratio of consumption to permanent income for groups of families at different levels of permanent income will be the same. The hypothesis in its extreme form states that the average fraction of permanent income devoted to consumption will be the same for groups of families with permanent income near the bottom of the income scale and for groups of families near the top of the income scale. In other words, the *apc* of families at all levels of family income is held to be the same when the *apc* is expressed as a ratio of permanent consumption to permanent income. This, of course, also means that the average propensity to save of families at all levels of family income is the same. The "rich" and the "poor" devote the same fraction of their incomes to saving!

This conclusion, which appears to conflict with what ordinary observation shows, follows from the argument of this theory that saving is primarily for the purpose of providing future consumption for the family. The family's intent is to even out consumption over a time period that is substantially longer than the single year but not necessarily as long as its remaining life span. Although this behavior is necessary if families at all income levels are to smooth out consumption, most economists question whether this is a correct description of actual behavior. It may be granted that most families make some attempt to even out consumption in this way, but it is questionable whether the preference for present over future goods is not greater for low-income families than it is for high-income families. However strong the desire of low-income families to avoid a level of consumption in later years that is even lower than the low level of the current year, it is difficult to believe that such families feel able to save the same fraction of their meager incomes that high-income families save of theirs. The pressures toward present consumption at the low-income levels are such that the preference for present goods over future goods would appear to be substantially stronger here than at the high-income levels. These pressures would be expected to operate to keep consumption by low-income families high relative to their incomes and to keep saving by high-income families high relative to their incomes.

Another basic argument of Friedman's permanent income theory is that the transitory component of consumption is not correlated with the transitory component of income. This amounts to saying that in a period in which a family's measured income contains a negative transitory component, it does not reduce its consumption in response, nor, under the opposite circumstance, does it raise its consumption in response. Unexpected increases or decreases in income thus result in equivalent increases or decreases in saving; consumption is unaffected by "windfall" gains or losses. In other words, the marginal propensity to consume out of transitory, or "windfall," income is held to be zero. As they have questioned the argument that the fraction of family income saved is the same at all levels of family income, most economists have questioned whether this too is a correct description of actual consumer behavior. In the words of one critic, the hypothesis says, "The man who has a lucky day at the races does not buy his friends a drink, and the poor fellow whose wallet is stolen does not postpone the purchase of a new overcoat."[10] This critic argues, and submits some evidence to suggest, that "the lucky winner does *not* run to the savings bank but to the tavern, and the victim of theft does cut his coat according to his cloth."[11]

However, to the degree that the marginal propensity to consume out of transitory income is indeed zero or at least very small, the permanent income theory carries a far-reaching impli-

[10] H. S. Houthakker, "The Permanent Income Hypothesis," in *American Economic Review*, June 1958, p. 398.
[11] *Ibid.*, p. 404.

cation, namely, that the marginal propensity to consume out of measured income may be quite unstable. In any time period, millions of persons realize changes in their measured incomes, and the view held by these persons as to the nature of that income change becomes a major determinant of the marginal propensity to consume out of measured income for all persons combined for that time period. The larger the portion of the income change perceived by its recipients to be transitory, the smaller the marginal propensity to consume, and vice versa. As one can hardly expect that the perception of those particular individuals whose incomes change will be the same from one period to the next, the marginal propensity to consume will not be stable from one period to the next. The existence of a relatively stable marginal propensity to consume was a major building block of Keynesian theory. The absolute income theory yielded a stable marginal propensity to consume, but the permanent income theory denies any such stability and as such has dealt a damaging blow to the Keynesian model. Still, the conclusion of an unstable marginal propensity to consume depends on the argument that the marginal propensity to consume out of transitory income is zero or at least very small, and, as noted, that argument is not without its critics.

The various basic arguments of the permanent income theory have raised considerable controversy and generated a sizable literature.[12] Empirical evidence has been presented on both sides, and the debate is far from resolved. Our purpose here is not to pursue this debate but merely to set forth some essentials of the permanent income theory and to relate these to the simple cross-sectional evidence provided by empirical family consumption functions.

Cross-sectional data such as those plotted in Figure 8-1 clearly do not form a pattern that would be expected on the basis of the permanent income theory. The data as given in the 1961 study, as well as data from similar studies made in other years, uniformly show the *apc* declining as we move from lower to higher levels of family income. However, this relationship emerges from data on the "current" income and the "current" consumption of families at different levels of income. The permanent income theory does not deny the relationship shown by such studies; what it does deny is the appropriateness of the measured income and consumption variables employed in such studies. According to this theory, the use of current or measured income improperly mixes together permanent and transitory income, on the one hand, and permanent and transitory consumption, on the other. The fact that the *apc* out of current income declines with higher levels of current family income is held to be attributable to the influence of the transitory income and consumption components in current income and current consumption. A family consumption function that relates permanent consumption to permanent income would look quite different—according to the theory, it would be a straight line from the origin, thus showing a constant proportion of income consumed at all levels of family income. Only by turning to cross-sectional estimates of perma-

[12] See, for example, I. Friend and I. B. Kravis, "Consumption Patterns and Permanent Income," in *American Economic Review*, May 1957, pp. 536–55; H. W. Watts, "Long-Run Income Expectations and Consumer Saving," in T. F. Dernberg, R. N. Rosett, and H. W. Watts, *Studies in Household Economic Behavior*, Yale Univ. Press, 1958, pp. 103–44; H. W. Watts, *An Analysis of the Effects of Transitory Income on Expenditure of Norwegian Households*, Cowles Foundation Discussion Paper 149, 1962; R. C. Bird and R. G. Bodkin, "The National Service Life Insurance Dividend of 1950 and Consumption: A Further Test of the 'Strict' Permanent Income Hypothesis," in *Journal of Political Economy*, Oct. 1965, pp. 499–515; R. Bodkin, "Windfall Income and Consumption," in *American Economic Review*, Sept. 1959, pp. 602–14; and M. E. Kreinen, "Windfall Income and Consumption," in *American Economic Review*, June 1961, pp. 388–90. See also T. Mayer, *Permanent Income, Wealth, and Consumption*, University of California Press, 1972, a book devoted to a synthesis of previous tests of the permanent income theory and the presentation of some new tests.

nent income and permanent consumption could one test empirically whether the *apc* out of permanent income is the same at all levels of family income. The tests that have been made offer conflicting results.[13] Simple cross-sectional data based on current income such as those given in Figure 8-1, however, do not directly shed light on this question.

Since the permanent income theory holds that the family consumption function relating permanent consumption to permanent income is a straight line from the origin, it leads directly to the conclusion that changes in aggregate permanent income give rise to proportional changes in aggregate permanent consumption. For, if the proportion of income saved is indeed the same at all levels of family income, changes in the permanent income of all families combined will have no effect on the proportion of aggregate income saved or consumed. This particular conclusion reached by the permanent income theory does not conflict with that reached by the relative income theory but it does conflict with that reached by the absolute income theory—although it must be noted that the conclusions are not strictly comparable, since the absolute and relative income theories are not expressed in terms of the same income and consumption concepts as the permanent income theory.

CONSUMER BEHAVIOR: MACROANALYSIS

APC CONSUMP/D. INC   MPC ΔCONS/ΔD. INC

Prior to World War II, empirical evidence on the consumption–income relationship at the macro level was limited to such evidence as could be drawn from family budget-study data. During World War II, however, estimates of national totals for personal consumption expenditures and disposable personal income became available for the first time on a comprehensive basis. Unlike cross-sectional data, these data show specifically how aggregate consumption expenditures have varied with aggregate disposable personal income.

These data for the years 1929–76, in constant (1972) dollars, are listed in columns (2) and (3) of Table 8-2. Column (4) shows the average propensity to consume, obtained by dividing the figures in column (3) by the figures in column (2). Column (5) shows the marginal propensity to consume, obtained by dividing the change in consumption from one year to the next by the change in disposable income for the same two years. Thus, in one simple table we have what appear to be all the data we need to discover how aggregate consumption has varied with aggregate income over a period of almost fifty years.

In Chapter 4 we assumed that this relationship has the following properties: (1) the MPC is positive but less than 1; (2) the MPC is the same for all changes in income; and (3) the APC decreases as income increases. How do these properties check out against the actual data? For the first, column (5) shows that in 15 of 47 cases the MPC is either negative, equal to, or greater than 1. For the second, there is considerable variability in the magnitude of the MPC; far from being the same for all annual changes in income, it seems to be different for every annual change in income. The third property is far from fully satisfied. In 13 of 48 cases, the APC rises with a rise in income, although the rise in many cases is so small that it may be

[13] Permanent income is, of course, not directly measurable. However, various ways of approximating permanent income on a cross-sectional basis have been devised and provide the basis for testing the theory. Because the theory holds that transitory consumption is not correlated with transitory income, measured consumption may be said to depend on permanent income. For some cross-sectional tests relating measured consumption to permanent income, see I. Friend and I. B. Kravis, *op. cit.*

accounted for by statistical error. In no case does it fall with a decline in income. Seven of the 13 exceptions are found in the ten-year period 1954–63. Does not this contrary evidence force us to reject the consumption–income relationship we earlier assumed, since it appears to be largely inconsistent with the way that consumers actually apportion their aggregate disposable income between consumption and saving?

These apparently pronounced differences, properly interpreted, do not necessarily lead to this conclusion. It was almost inevitable that there would be differences, even pronounced differences for some years, between the consumption–income relationship we adopted in developing the simple theory of income determination and the one revealed by the aggregate data. The reason is that the former is a relationship between income and consumption drawn on the assumption that all the factors that influence consumption, other than

TABLE 8-2
Disposable personal income and personal consumption expenditures (billions of 1972 dollars)

(1) Year	(2) Disposable Personal Income	(3) Personal Consumption Expenditures	(4) APC	(5) MPC	(1) Year	(2) Disposable Personal Income	(3) Personal Consumption Expenditures	(4) APC	(5) MPC
1929	229.8	215.6	0.938		1953	397.5	364.2	0.916	0.864
1930	210.6	200.0	0.950	0.812	1954	402.1	370.9	0.922	1.457
1931	201.7	192.1	0.952	0.888	1955	425.9	395.1	0.928	1.017
1932	174.3	174.1	0.999	0.657	1956	444.9	406.3	0.913	0.589
1933	169.7	170.7	1.006	0.739	1957	453.9	414.7	0.914	0.933
1934	179.7	177.2	0.986	0.650	1958	459.0	419.0	0.913	0.843
1935	196.6	188.1	0.957	0.645	1959	477.4	441.5	0.925	1.223
1936	220.7	206.8	0.937	0.776	1960	487.3	453.0	0.930	1.162
1937	227.8	214.3	0.941	1.056	1961	500.6	462.2	0.923	0.692
1938	212.8	209.2	0.983	0.340	1962	521.6	482.9	0.926	0.986
1939	230.1	220.3	0.957	0.642	1963	539.2	501.4	0.930	1.051
1940	244.3	230.4	0.943	0.711	1964	577.3	528.7	0.916	0.717
1941	278.1	244.1	0.878	0.405	1965	612.4	558.1	0.911	0.838
1942	317.3	241.7	0.762	−0.061	1966	643.6	586.1	0.911	0.897
1943	332.2	248.7	0.749	0.470	1967	669.8	603.2	0.901	0.653
1944	343.9	255.7	0.744	0.598	1968	695.2	633.4	0.911	1.189
1945	338.6	271.4	0.802	−2.962	1969	712.3	655.4	0.920	1.287
1946	332.4	301.4	0.907	−4.839	1970	741.6	668.9	0.902	0.461
1947	318.8	306.2	0.960	−0.353	1971	769.0	691.9	0.900	0.839
1948	335.5	312.8	0.932	0.395	1972	801.3	733.0	0.915	1.272
1949	336.1	320.0	0.952	12.000	1973	854.7	767.7	0.898	0.650
1950	361.9	338.1	0.934	0.702	1974	842.0	760.7	0.903	0.551
1951	371.6	342.3	0.921	0.433	1975	857.3	775.1	0.904	0.941
1952	382.1	350.9	0.918	0.819	1976	890.3	821.3	0.922	1.400

SOURCE: *The National Income and Product Accounts of the United States, 1929–74, Statistical Tables,* and *Survey of Current Business,* July 1977, U.S. Department of Commerce.

c $\Delta C/\Delta Y$

C_a

the level of income, remain unchanged; the latter, on the other hand, is based on the measured year-to-year values for consumption, which are what they are as a result of all the factors that influence consumption. In other words, the changes in actual consumption are certainly influenced by the changes in income, but there are other factors at work such as the price level, interest rates, income distribution, consumer asset holdings, and consumer credit terms. These factors also change and in so doing exert some influence on the level of consumption spending from year to year.

Ideally what is needed is a way of isolating the portion of the actual changes in consumption that were due to changes in income from the portion that were due to the changes in all the other factors that influence consumption. Although advanced statistical techniques enable us to derive quantitative approximations of this kind, for present purposes a simpler approach will suffice.

The data from columns (2) and (3) of Table 8-2 are plotted in Figure 8-3. Each dot on the chart indicates the combination of personal consumption expenditures and disposable personal income for a given year between 1929 and 1976. A glance at the scatter of the dots shows that consumption does not vary with income exactly in the way suggested in Chapter 4. To do so, a straight line with its intercept above zero on the vertical axis and with a positive slope of less than 1 would have to pass through every dot in the figure. Although no straight line will do this, there is one that comes close to doing it if we ignore the six dots for the years 1941–46. Omitting these six years, the straight line drawn in the chart is the straight line fitted to the data for 1929–40 and 1947–76 by the method of least squares.[14] The equation for this empirical consumption function is $C = 15.9 + 0.89\,Y$.[15] The vertical differences between the dots and this straight line are quite apparent. However, in view of the number of other factors that influence consumption spending, what is surprising is not that we find such differences but how small these differences actually are. The relationship between disposable income and consumption expenditures is remarkably close and seems to suggest that consumption is very largely explained by income. Since consumption so closely follows income, the many other factors that influence consumption spending, taken as a group, would seem to exert a far smaller influence on consumption spending than does the single factor of the level of disposable income.[16]

[14]This is the straight line of "best fit" in the sense that the sum of the squares of the deviations between actual consumption and consumption indicated by the line, known as a *regression line* or simply as a *regression*, will be less for this straight line than for any other that could be fitted to the data.

[15]This equation is the familiar one for the consumption function used in Chapter 4 ($C = C_a + cY$), only now, instead of hypothetical figures for C_a and c, actual values computed from the data for the United States for 1929–40 and 1947–76 yield the empirical consumption function for this period. C_a, which is here \$15.9 billion, indicates the height of the consumption function, or the point at which it intercepts the vertical axis. As earlier described, it is the amount of consumption spending at the zero level of income, recognizing that in reality the level of income could not fall to zero; c, which is here 0.89, is the MPC (or $\Delta C/\Delta Y$ with Y being disposable personal income), or the slope of the consumption function. If consumption were to change with every change in income as given by the equation, then for every change of \$100 million in income, consumption would change by \$89 million.

Actual empirical work on the consumption function involves much more complex functions than the simple equation in one variable used here. For an introduction to empirical estimates of the consumption function, see M. K. Evans, *Macroeconomic Activity*, Harper & Row, 1969, Ch. 3, and the references given there.

[16]Conclusions such as these, suggested by the single-equation, least-squares method used here, must be interpreted with great caution. This statistical method is acceptable for dealing with variables that are independent, but the consumption and income variables are interdependent. Consumption depends on income, and income also depends on consumption. Therefore, we cannot conclude that the changes in income are the causes of changes in consumption from the fact that a line of regression shows a close relationship of consumption to income. The line of causation could run in the other direction, changes in income being caused by changes in consumption. What we have is a relationship of mutual interaction in which higher income means higher consumption and higher consumption means higher income. Furthermore, when we note that consumption expenditures typically constitute about 90 percent of income, it is apparent that we would find a close relationship between income and consumption even if con-

A glaring exception to this conclusion is the period 1941–45 and perhaps also 1946, when the otherwise close relationship between consumption and disposable income seems to have broken down completely. The unusual distance of the dots for these years from the regression line is readily explained by the extraordinary impact of certain nonincome factors

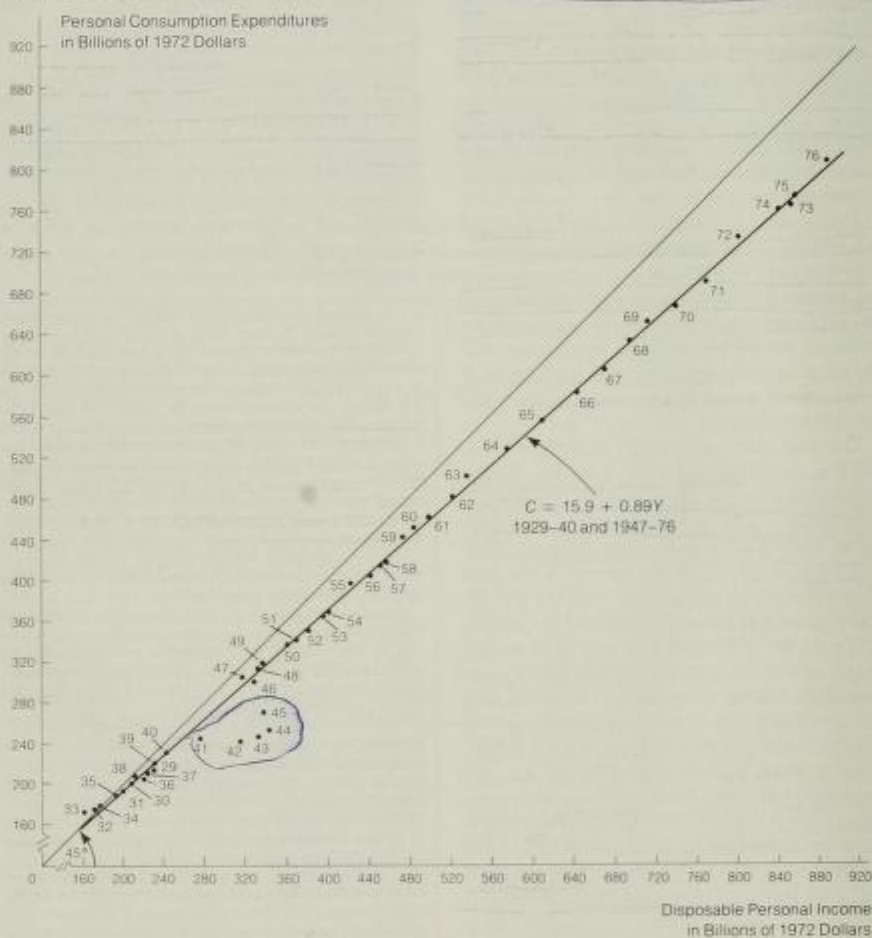


FIGURE 8-3
Relationship between personal consumption expenditures and disposable personal income, 1929–1940 and 1947–1976

in operation during World War II.[17] It is statistically proper to disregard these years in determining the "normal" consumption–income line that characterizes the relationship between income and consumption for a period of years other than those of a major war. Dropping 1941–46 for this reason, the remaining years may be somewhat arbitrarily divided into four periods of approximately equal length: 1929–40, 1947–56, 1957–1966, and 1967–1976. A separate regression line fitted to the data for each of these periods will shed more light on the consumption–income relationship, for we can see how the fit of the line for each of these shorter periods compares with that for the overall period and how the shorter periods compare with one another.[18] The lines for the four shorter periods are shown in Figure 8-4, where the equation for the 1929–40 line is $C = 33.9 + 0.80Y$; for the 1947–56 line, $C = 41.2 + 0.82Y$; for the 1957–66 line, $C = 15.3 + 0.89Y$; and for the 1967–76 line, $C = 8.0 + 0.90Y$.

In comparing each of these regression lines with that in Figure 8-3, which reflects the overall period, it is apparent that the regression lines for the shorter periods provide a better fit to the years on which each is based. It is, of course, still obvious that there are factors other than income influencing consumption during each of these four periods. In no year, however, did these factors produce a level of consumption spending that was widely different from that which would be expected on the basis of the regression line fitted to the data of the relevant period.

It is also apparent in Figure 8-4 that the re-

---

sumption were completely unrelated to income. D. B. Suits has indicated that the correlation between consumption and income in these circumstances may be expected to be about 0.9 even if income changes do not cause consumption changes. See D. B. Suits, "The Determinants of Consumer Expenditure: A Review of Present Knowledge," *Impacts of Monetary Policy*, Commission on Money and Credit, Prentice-Hall, 1963.

Where interdependence exists between variables, the customary terminology of independent variable (income) and dependent variable (consumption) and the single-equation, least-squares method of deriving the relationship between the two are inappropriate. Each variable should have an equation of its own to explain the way it varies in a system of equations that are solved simultaneously. This is the method most frequently used today by econometricians in consumption-function estimation. For the introductory purpose of this chapter, we will limit ourselves to the simple one-equation, least-squares method, but it is essential to point out that this method is inherently defective in handling interdependent variables and that conclusions reached by means of it cannot be taken at face value. For a discussion of this problem, see Evans, *op. cit.*, pp. 48–55.

[17]Consumers spent much less than would have been expected solely on the basis of their incomes during 1941–45 because of the complete unavailability of certain durable goods like automobiles and the limited availability of other goods. To a lesser degree, consumers also spent less than would have been

to save and p

tered just above, just below, or right along this extended broken line. Instead, the dots all lie above the extended line. Apparently, in the early postwar period, consumers chose a permanently higher level of consumption relative to income than they chose in the prewar period. The same relationship is found between the 1957–66 and 1947–56 periods and between

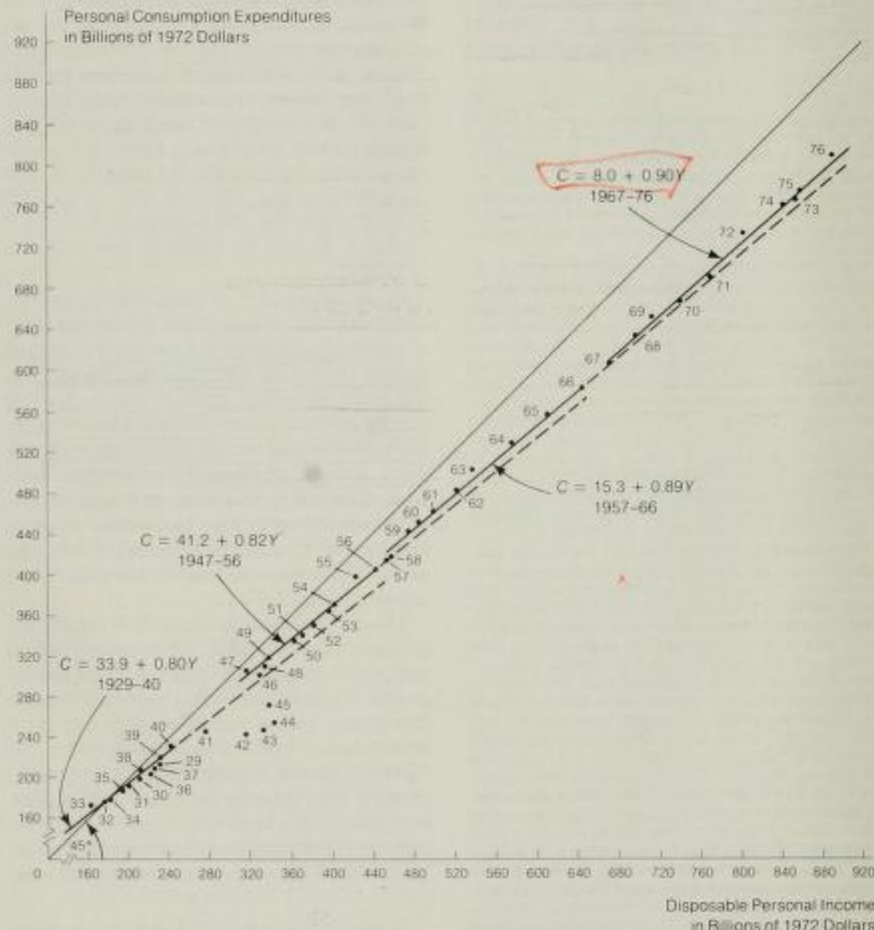


FIGURE 8-4
Relationship between personal consumption expenditures and disposable personal income, 1929–1940, 1947–1956, 1957–1966 and 1967–1976

the 1967–76 and 1957–66 periods, but to a less noticeable degree.[19]

The purpose of this examination of the time series data has been to see what support, if any, the data give to the kind of theoretical consumption function we adopted in Chapter 4. When we look at the data on a year-to-year basis as a preliminary step, it is true that we find little support: The MPC varies considerably, at times even exceeding 1 or falling below zero, and the APC does not always fall with a rise in income. However, when we fit a linear regression line that "averages" the data for a number of years, we find that the line derived gives a good fit to the observations. If such a good fit is provided by a straight line with a positive slope less than 1 and an intercept with the vertical axis above zero, it follows, on the average for the period covered, that the APC decreases with rises in income and that the MPC is positive but less than 1, the major assumptions underlying the earlier theoretical consumption function. At this point at least, the empirical evidence seems to give substantial support to the type of theoretical consumption function we employed in Chapter 4.

CYCLICAL AND SECULAR CONSUMPTION—INCOME RELATIONSHIPS

The regression lines, or empirical consumption functions, derived from the time series data for 1929–40, 1947–56, 1957–66, and 1967–76 may be described as *cyclical consumption functions*. Although each period covers more than a single business cycle, as that term is now most commonly understood, we may designate them as cyclical to distinguish between a period of such limited duration that the changes in consumption spending may be related to the cyclical ups and downs in disposable income and a much longer period, running over many decades, in which the changes in consumption spending may be related to the secular growth in disposable income. In such a long period, we may abstract from the cyclical fluctuations and derive what may be described as the *secular consumption function*. We may also describe the cyclical consumption function as the short-run consumption function and the secular as the long-run.

In the previous section, we saw that the theoretical consumption function adopted in Chapter 4 received support from the aggregate time series data. It may be asked why this does not end the matter. Why now cyclical and secular consumption functions? In a formal sense, the answer is that the empirical consumption function revealed by the data for a long-run period is different from that revealed by the data for a short-run period. What is the basic relationship between disposable income and consumption if all factors that influence consumption—other than income—are absent? If the time series data suggest a different relationship according to the time interval examined, our question remains unanswered until we explain this difference.

Figure 8-4 has already shown the nature of the problem. The regression line for 1929–40 gives a good fit to the data for that short-run period, and the regression lines for 1947–56, 1957–66, and 1967–76 give good fits to the data for those periods. In each case, the MPC is positive but less than 1 and is less than the APC—the properties we assumed in the theo-

[19] There are no years omitted between the postwar periods as there are between the prewar and first postwar periods. We find that income in the first year of the second postwar period, 1957, is only $9 billion above the income of the last year of the first postwar period, 1956, whereas income of the first year of the first postwar period, 1947, is almost $75 billion above income of the last year of the prewar period, 1940. Other factors are involved, but this one in part accounts for the relatively smaller upward shifts in the regression lines for the last two postwar periods.

retical consumption function of Chapter 4. Although this much is apparent, a problem arises from the fact that the line for each later period lies above the line we get by simply projecting the line for the preceding period. This suggests that the 1929–40 line may in turn have shifted upward from a regression line for a preceding short-run period, say 1919–28, and that a regression line to be computed for 1977–86 when that period is behind us will also have shifted upward from the 1967–76 line. If this indeed is what takes place over the long run, the argument that the APC decreases with a rise in income is supported by the data *only* if we limit the application of that argument to a description of the short-run relationship between disposable income and consumption. Only then does the APC plainly decrease with increases in income. On the other hand, if from one short-run period to the next the entire function shifts upward, it does not necessarily follow that the APC will decrease with long-run increases in income. *The upward shift in the function could be such as to produce an APC that remains roughly unchanged in the long run.*

In fact, the data suggest that this is what actually does happen over the long run. To demonstrate this empirical long-run relationship, the data of Table 8-2 will not suffice. First, the period covered is not long enough. Second, annual data reflect changes within the short run (intracycle) rather than changes between successive short-run periods (intercycle). Table 8-3 gives us the kind of data we need for our purposes. The time span covered is seventy years, and the estimates for income in column (2) and for consumption in column (3) cover the full decades (here overlapping) that are shown in column (1), so that the changes in the figures are changes between successive short-run periods.[20] The APC's in column (4) in this case represent the ratio of the estimated total of consumption expenditures for each ten-year period to the estimated total of national income for each corresponding ten-year period.

Table 8-3 shows us that although national income increased more than sevenfold from the 1869–78 period to the 1919–28 period, the APC varied only over the narrow range of 0.84 to 0.89. The last two periods, 1924–33 and 1929–38, both show an APC that is higher than the APC for any earlier period, the reason being that both these periods include the years of the Great Depression. Whether or not we include these last two periods, it is significant that the results listed in column (4) are quite different from those given by the time series data for the short-run periods examined earlier. In the short run, we found that as the level of income rose, the APC tended to decline; now we find that in the long run, as the level of income rises, the APC remains quite stable. Instead of finding

TABLE 8-3
National income and personal consumption expenditures, 1869–1938 (billions of 1929 dollars)

(1) Decade	(2) National Income	(3) Personal Consumption Expenditures	(4) APC
1869–78	9.3	8.1	0.86
1874–83	13.6	11.6	0.86
1879–88	17.9	15.3	0.85
1884–93	21.0	17.7	0.84
1889–98	24.2	20.2	0.84
1894–1903	29.8	25.4	0.85
1899–1908	37.3	32.3	0.86
1904–13	45.0	39.1	0.87
1909–18	50.6	44.0	0.87
1914–23	57.3	50.7	0.89
1919–28	69.0	62.0	0.89
1924–33	73.3	68.9	0.94
1929–38	72.0	71.0	0.99

SOURCE: Columns (2) and (3) are from Simon Kuznets, *National Product Since 1869*, National Bureau of Economic Research, Princeton Univ. Press, 1946, p. 119, and column (4) is from Simon Kuznets, *National Income: A Summary of Findings*, National Bureau of Economic Research, Princeton Univ. Press, 1946, p. 53.

[20] Note that Table 8-3 shows national income rather than disposable personal income. For these early years, no satisfactory estimates of disposable personal income are available. Also, for the decades 1869–78 through 1914–23, estimates are for decades; and for the decades 1919–28, 1924–33, and 1929–38, estimates are averages of annual estimates.

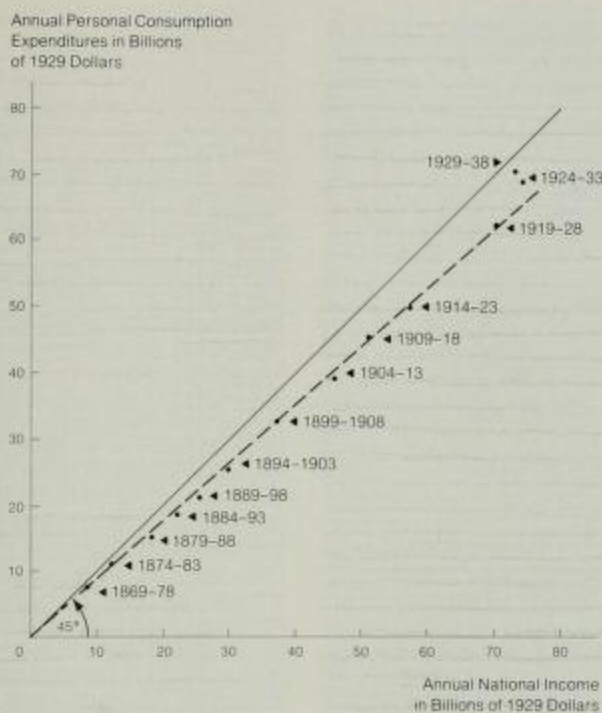


FIGURE 8-5
Secular consumption function

that an ever smaller fraction of income was devoted to consumption as the level of income doubled and redoubled over the decades, we find that an approximately stable proportion of income was devoted to consumption decade after decade.[21]

The differences between the cyclical and secular relationships become more apparent when the data of columns (2) and (3) in Table 8-3 are plotted, as in Figure 8-5. Apart from the dots for the 1924–33 and 1929–38 periods, both of which were pushed upward by the extraordinary severity of the Great Depression, a visually fitted straight-line consumption func-

[21] The findings of another study agree with those summarized in Table 8-3. See R. Goldsmith, *A Study of Saving in the United States*, Vol. 1, Princeton Univ. Press, 1955, pp. 393, 400.

Although the absolute magnitude of the ratio is different due to differences in definition, the ratio in the postwar period shows the same stability as the ratio in Table 8-3. For the overlapping periods 1947–56, 1952–61, 1957–66, 1962–71, and 1967–76, the ratio of personal consumption expenditures to national income, both in 1972 dollars, was 0.76, 0.74, 0.75, 0.74, and 0.77, respectively. For these same periods, the ratio of personal consumption expenditures to disposable personal income, both in 1972 dollars, was 0.93, 0.92, 0.92, 0.91, and 0.91, respectively.

tion drawn *from the origin* comes very close to passing through all the other dots in the figure. Unlike the general equation $C = C_a + cY$ for the *cyclical* straight-line consumption function that intersects the vertical axis above zero, the general equation for the *secular* straight-line consumption function is simply $C = bY$, in which the long-run MPC is designated by b to distinguish it from the short-run MPC, which is designated by c. The C_a constant becomes zero in the long-run function; the straight line intercepts the vertical axis at the origin. Since the APC equals the MPC at all levels of income for any straight-line consumption function that intersects the vertical axis at the origin, the long-run consumption function is one in which APC = MPC at all levels of income, in contrast to the short-run function in which APC > MPC at all levels of income. Another way of distinguishing the two functions is to describe the short-run straight-line function as *nonproportional* and the long-run straight-line function as *proportional*. This simply refers to the fact that, in the short run, consumption does not change proportionally with income; this proportion instead rises with falling income and falls with rising income. In the long run, consumption changes proportionally with income; it remains roughly the same proportion of income as the level of income doubles and redoubles over the decades that make up the long run.

On the basis of the empirical consumption functions given in Figures 8-4 and 8-5, we may now construct Figure 8-6 to show in a general way the nature of the short-run and long-run functions suggested by the time series data. Below the 45° line lies a line, labeled *LR*, that depicts the proportional relationship between disposable income and consumption suggested by the data for the long-run period. Each of the family of lines labeled *SR* represents the nonproportional relationship between disposable income and consumption suggested by the data for a short-run period. Figure 8-4 showed that the short-run consumption function shifted upward from the 1929–40 period to the 1947–56 period, from the 1947–56 period to the 1957–66 period, and again from the 1957–66 period to the 1967–76 period. Each successively higher *SR* line in Figure 8-6 represents such a shift in the function from one short-run period to the next. To understand the figure, it is essential to keep the time dimensions in mind. The *LR* curve relates to data covering an unbroken series of consecutive decades and designates the movement of consumption in relation to the movement of income decade by decade; each *SR* curve relates to only a segment of this long-run period and designates the movement of consumption in relation to the movement of income year by year over this segment of the long-run period.

RECONCILING THE CYCLICAL AND SECULAR RELATIONSHIPS: THE ABSOLUTE, RELATIVE, AND PERMANENT INCOME THEORIES

What is the "true" or underlying relationship between consumption and disposable personal income? If all other factors influencing consumption were absent, would consumption vary nonproportionally (along a curve such as *SR*) or proportionally (along a curve such as *LR*) as income changed? In other words, is the relationship basically nonproportional, as depicted by a consumption function such as *SR*, but with this function shifting upward over time at just the rate that produces a long-run proportional relationship? Or is the basic relationship one in which consumers as a group increase their consumption proportionally with the change in income, as shown by a consumption function like *LR*, with the nonproportional functions like *SR* indicating departures from the basic proportional relationship?

Relative theories

The absolute income theory takes the former position; the relative and permanent income theories take the latter position. We may see how the several theories lead to these positions as the theory of individual behavior of each is generalized to cover the aggregate behavior that lies behind the time series data.

The Absolute Income Theory

The basic tenet of this theory is that the individual consumer's spending depends on the absolute level of his income. We saw earlier that when this theory is extended to aggregate behavior, it indicates that increases in the aggregate income of all consumers should result in a decline in the APC, or the fraction of this aggregate income devoted to consumption. Since the actual level of aggregate income grows larger and larger in the long run, this theory leads to the conclusion that the APC should become smaller and smaller.[22] In terms of Figure 8-6, as income increases in the long run, the theory would lead us to expect consumption to follow the path of an *SR* curve projected out to the right. However, the data for the long run show clearly that this is not what happens. One way of reconciling the position of the absolute income theory that the basic relationship is nonproportional with the data that show the long-run relationship to be proportional is through an upward shift in the basic, nonproportional consumption function as a result of changes in factors other than income. There are a number of factors working in this direction: (1) With the increase in the accumulated wealth of households that has accompanied the long-run growth of income, households have tended to spend a larger fraction of any given level of income, thus contributing to an upward shift in the consumption function. (2) Over the present long-run period, there has been a continuing movement of population from the farms to the cities. Since the propensity of urban wage earners to consume is substantially higher than that of farm proprietors, this shift in population has contributed to an upward shift in the consumption function. (3) The percentage of older people in the population has increased over the long-run period. Since per capita consumption does not drop off as rapidly as does per capita income of this age group, the consumption function tends to shift upward as this age group becomes a larger part of the population. (4) New consumer goods have been introduced at a rapid rate over this long-run period. As more and more of these goods come to be regarded as "essentials" by the typical household, the consumption function tends to shift upward.

Factors like these, according to the absolute income theory, have caused the consumption function to shift upward by roughly the amount necessary to produce a proportional relationship between consumption and income over the long run and thus to prevent the appearance of what would otherwise be the nonproportional relationship that would be expected on the basis of the income factor alone.

The Relative Income Theory

Following the relative income theory, we reach the opposite position—that the basic relationship between consumption and income is proportional or is such that consumption in the long run tends to follow a path like that of the *LR* curve in Figure 8-6. The relative income theory must accordingly explain why the short-run relationship departs from proportionality.

When this theory, as earlier considered for individual consumer behavior, is extended to aggregate behavior, it is transformed into one

[22]When the period under consideration stretches out over decades as it does here, it is essential to express the consumption and income variables in per capita form. If aggregate real income grows no faster than total population, real income per capita remains unchanged, and there is no reason to expect a change in the APC on the basis of income alone. The situation that is relevant here is that in which real income per capita grows larger and larger, and in this situation the absolute income theory holds that the ratio of real consumption per capita to real income per capita will become smaller and smaller.

(1) accumulation of wealth
(2) Population shift
(3) ↑ in % of old people whose I + C –
(4) New consumer goods considered essent.

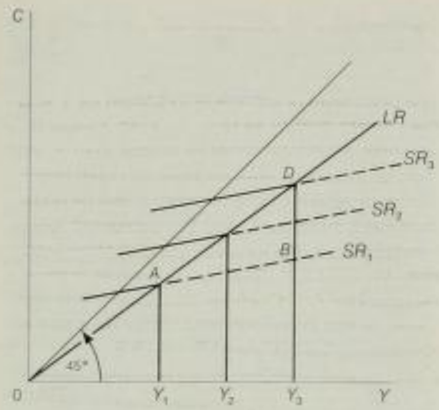


FIGURE 8-6
Secular and cyclical consumption functions

that makes the ratio of consumption to income depend on the ratio of the *current income level* to the *peak income level* previously attained. In the aggregate case, consumption in the current period thus depends on income in that period *relative* to the peak income level previously attained. In developing the peak income argument of the relative income theory, Duesenberry takes as his point of departure the known fact that income does not grow at an even pace over the long run but instead displays spurts and dips known as business cycles. These cyclical ups and downs in the level of income produce the nonproportional consumption–income relationship found in the short run. If there were no short-run income fluctuations of this sort, consumption would follow an uninterrupted path such as that traced out by the *LR* curve. However, there are such fluctuations in income that push the economy off the *LR* and onto the successive *SR* curves in an unbroken series over time.

To illustrate Duesenberry's theory in terms of Figure 8-6, let us suppose that a recession strikes when the economy is at income Y_1, which we will assume represents the highest level of income yet reached by the economy. As income falls over the course of the recession, consumers as a group will attempt to maintain the consumption level they enjoyed when income was at its peak. Consumers strongly resist any reduction in their standard of living but eventually find it necessary to cut back their expenditures gradually as their aggregate income falls. Thus, consumption falls but not proportionally with the fall in income. The economy moves down the SR_1 curve, and as it does the APC rises and the APS falls. This reflects the fact that in an attempt to maintain the previously established higher standard of living, consumers will cut down their saving more than proportionally and consumption less than proportionally with the fall in income. Then, as recovery succeeds recession and the income level begins its move back toward its previous peak, consumption rises but less than proportionally with the rise in income. Consequently, the APC falls and the APS rises. The movement of consumption upward along the SR_1 curve is just the reverse of the earlier movement downward along this curve. Just as consumers cut down saving more than proportionally to defend the level of consumption on the downswing, so they increase saving more than proportionally on the upswing in order to restore saving to its previous level.

As the level of income rises to its previous peak and pushes on to higher ground, consumption does not follow the SR_1 line to the right past the peak income but instead shifts over to and follows the *LR* line. Since income is now rising to levels not previously attained, consumers no longer feel the increased urge to save that they felt during the period of recovery. They are now enjoying higher levels of income and feel better off and more certain of the future. They feel free to consume a larger fraction of this income than they consumed of the below-peak income of the preceding recession. As income grows to ever higher peaks, consumption increases proportionally; the APC

and APS remain unchanged. If income should grow at a steady, uninterrupted rate over time, consumption would follow the proportional LR curve upward without varying. However, since it is the nature of the economy that a business downturn will succeed an upturn just as an upturn will succeed a downturn, instead of an uninterrupted movement along LR, the economy will slide into another recession after having established a new peak income of Y_2. The sequence is now a repetition of that just described but with the sequence traced out along the SR_2 curve from the peak income of Y_2. Recession will cause consumption to move down the SR_2 curve; recovery will cause it to move up the SR_2 curve until the previous peak at Y_2 is reattained. At this point consumption again will shift over to the LR curve, and the economy will move ahead to a new peak at Y_3.[23]

The aggregate behavior suggested by the relative income theory may be expressed in a single consumption function that combines the properties of the long-run function, $C = bY$, and the short-run function, $C = C_a + cY$, as follows:

$$C = \overline{Y}(b - c) + cY$$

The peak level of income is indicated by $\overline{Y}$ and the current level of income by Y, so Y must for any time period be equal to or less than $\overline{Y}$. As long as income is steadily attaining new peaks during prosperity, peak income, $\overline{Y}$, will be equal to current income, Y, and the combined function reduces to the long-run consumption function $C = bY$.[24] The consumption–income relationship described by this function is the same as that described graphically by a movement along the LR line of Figure 8-6.

If a recession strikes and the current level of income drops below the peak attained earlier, $\overline{Y}$ is constant and greater than Y. With $\overline{Y}$ constant, $\overline{Y}(b - c)$ is also constant, so the constant C_a may be substituted for it. With current income below peak income, the combined function thus reduces to the short-run consumption function $C = C_a + cY$. As long as current income is below peak income, the consumption–income relationship described by this function is the same as that described graphically by a movement along an SR line of Figure 8-6.

With recovery and the expansion of income beyond the previous peak, the combined function again reduces to the long-run function, $C = bY$, and consumption again increases proportionally with income along the LR line.

The Permanent Income Theory

This theory, like the relative income theory, holds that the basic relationship between consumption and income is proportional, but the relationship here is between permanent consumption and permanent income. The figures given in Table 8-3 show long-run proportionality between consumption and income, with these quantities being measured quantities; dis-

[23] Although we earlier described such periods as 1929–40, 1947–56, 1957–66, and 1967–76 as cyclical periods, it was then noted that each such period is actually made up of more than one complete business cycle. Thus, over the years 1948–75, GNP in constant dollars moved from peak, P, to trough, T, to peak, P, as follows (Roman numerals indicate quarters): P—1948 IV, T—1949 IV, P—1953 II, T—1954 II, P—1957 III, T—1958 I, P—1960 I, T—1960 IV, P—1969 III, T—1970 IV, P—1973 IV, and T—1975 I. Whether we measure from peak to peak or from trough to trough, each intervening period is a full business cycle, if we simply date the turning points of the cycle by the peaks and troughs in real GNP. It is to such cycles that the SR curves of Figure 8-6 most meaningfully apply. If one examines the consumption and disposable income data quarter by quarter for each of the full cycles during 1948–75, one finds a marked tendency for the APC to rise and the APS to fall quarter by quarter during business cycle contractions and vice versa during expansions, which is consistent with the nonproportional SR curves of Figure 8-6. However, this is a tendency and not a rule. An exception is found in the recession from 1969 III to 1970 IV during which the APS rose from 0.086 to 0.104 as the recession continued. During the expansion from 1970 IV to 1973 IV, the APS fell from 0.104 to 0.082 in 1972 III but then rose to 0.112 by the last quarter of that expansion. For the most recent contraction, 1973 IV to 1975 I, the APS changed in line with the tendency noted; it fell from 0.112 in 1973 IV to 0.089 in 1975 I.

[24] $C = b\overline{Y} - c\overline{Y} + cY$. As long as $\overline{Y} = Y$, $-c\overline{Y} + cY = 0$, and $C = b\overline{Y}$ or bY.

covering what the empirical long-run relationship between the corresponding permanent quantities may be requires that figures be obtained for these quantities.

Because the concept of permanent income inherently involves expected or anticipated income, there is no way of estimating it in any direct manner. What Professor Friedman does in his time series analysis is to take permanent income of any one year to be a weighted average of actual or measured income over a 17-year period. Measured income of the current year is weighted at 33 percent, that of the preceding year at 22 percent, and so forth for the other fifteen years, with the weights rapidly declining. On the basis of such estimates of permanent income for each year back to the early part of the century, Friedman finds the ratio of permanent consumption to permanent income approximately constant year after year. Excluding the war years, the relationship for 1905–51 is $C = 0.88Y$ (in which the variables refer to permanent quantities). The average propensity to consume equals the marginal propensity to consume and the function corresponds to a line like *LR* in Figure 8-6 with a specific slope of 0.88. We thus find a long-run relationship of proportionality whether the relationship is derived as earlier described from figures of the decennial measured kind given in Table 8-3 or from estimates of annual permanent consumption and permanent income as in Friedman's work.

The problem of reconciling the short-run nonproportional relationship with the long-run proportional relationship encountered in the cases of the absolute and relative income theories does not appear in the present case. As discussed earlier, the permanent income theory argues that the ratio of permanent consumption to permanent income is the same at all levels of family income, and this argument is directly translatable into a proportional relationship between aggregate permanent consumption and aggregate permanent income. In other words, if the ratio of permanent consumption to permanent income is the same at all levels of family income, this ratio in aggregate terms will remain the same as aggregate permanent income rises over the years. If there were a way of ascertaining what part of each year's aggregate measured income was permanent and what part transitory, we would have time series data that could provide a quite precise test of the relationship in question. However, no such exact separation is possible on an individual year basis—strictly speaking, this would require annual data on how much of his income each person views as permanent and how much as transitory and these data would have to be available year after year. Clearly, nothing resembling such data is at hand. Although it is recognized that the fraction of any one year's measured income that is permanent will differ from the same fraction for the preceding or following year, we see that there is no exact way of taking this into account on an individual year basis. Friedman has chosen the approach noted earlier, the use of a 17-year weighted average, as one that may provide a reasonable approximation of what part of each year's income is permanent and what part transitory.

Since that approach assigns a weight of 33 percent to the current year, 33 percent of the measured income of the current year is regarded as permanent income.[25] Therefore, if the income of the current year should fall $10 billion below that of the preceding year, the decrease in the current year's permanent income would be estimated at $3.3 billion. From

[25] Note that this does not say that the current year's permanent income is 33 percent of its measured income. The current year's permanent income also includes 22 percent of the preceding year's measured income and portions of the measured income of the preceding fifteen years. When these amounts are summed, the permanent income of the current year may exceed its measured income. It will be seen that this will happen if there is a sufficient decline in the current year's measured income relative to that of the preceding years. In the same way, the permanent income of the current year may be less than its measured income, a result which will follow if there is a sufficient rise in the current year's measured income relative to that of the preceding years.

the long-run estimate that the marginal and average propensity to consume out of permanent income are both 0.88 (and out of transitory income, zero), the change in consumption in the illustration would be $0.88 \times 0.33 \times \$10$ billion or approximately \$2.9 billion. This says that the "one-year" marginal propensity to consume is about 29 percent.[26] At first glance, this also seems to describe a situation graphically portrayed by a movement along a nonproportional consumption function like SR_1 (or SR_2 or SR_3) in Figure 8-6 in comparison to a movement along a proportional consumption function like *LR* in the same figure. However, to draw such a conclusion involves a variation of the error warned against on p. 118 in connection with cross-sectional data. If we want to use Figure 8-6 in connection with the permanent income hypothesis, we are required to plot permanent income, not measured income, along the horizontal axis and permanent consumption, not measured consumption, on the vertical axis. Therefore, if a \$10 billion decrease in measured income involves a decrease of \$3.3 billion in permanent income, and if this decline in permanent income involves a \$2.9 billion decrease in consumption, we move \$3.3 billion leftward on the horizontal axis and \$2.9 billion downward on the vertical axis, or we move along a proportional curve with a marginal and average propensity to consume of 0.88. Assuming that *LR* in Figure 8-6 had a slope of 0.88, this is the curve we would move down.

As long as we are dealing with permanent income and permanent consumption, the argument of the permanent income hypothesis is that the response of consumption to income will be a proportional one or one of the kind described by a movement along a curve like *LR* in Figure 8-6. The nonproportional *SR* curves of that figure do not exist in this case because the response of consumption to permanent income is the same in the short run as in the long run. There is no problem of reconciling the results found in the short run and in the long run as there was for the relative and absolute income theories, because here the response is proportional.

INCOME AND CONSUMPTION—A CONCLUDING NOTE

Empirical data covering well over a half-century tell us that the long-run relationship between consumption and income has been approximately proportional. We have seen that the relative and permanent income theories hold that this is the basic, underlying relationship between consumption and income, whereas the absolute income theory holds that the basic relationship is nonproportional. Which of the two is the "true" relationship is not a question of mere academic interest. If consumption is not basically proportional to income, we may expect that an ever larger fraction of income will be saved as income grows over the long run, unless this tendency toward a rising APS (or declining APC) is offset by recurrent upward shifts in the consumption function itself, caused by such factors as the accumulating wealth of households, the growing urbanization of the population, the shifting age composition of the population, and the appearance of more and

[26]As will be discussed briefly in Chapter 24, this aspect of the permanent income theory has a striking implication for the efficacy of fiscal policy. An increase in personal income taxation designed to reduce consumer expenditures will have a relatively small first-year effect on those expenditures, as the reduction in permanent income is only a fraction of the reduction in measured income. In terms of the illustration above, a decrease in disposable personal income of \$10 billion brought about by a \$10 billion increase in personal income taxes will reduce consumption by only \$2.9 billion in the first year. The marginal propensity to consume out of measured income is only 0.29. This is a far cry from the result suggested by the Keynesian absolute income hypothesis in which the reduction in consumption would be in the area of \$9 billion because the marginal propensity to consume out of measured income, in terms of that theory, is about 0.9.

more new consumer goods that come to be regarded as "essentials" in the standard of living. Although some of these factors may continue to exert the same influence over the indefinite future, we cannot be certain that the influence of all such factors combined will be just enough to shift the consumption function upward at the rate necessary to give long-run proportionality between income and consumption.

The implications of such a development in terms of the simple theory of income determination are apparent. If the basic consumption function is nonproportional, then the higher the level of income grows, the greater the percentage of that income that will be devoted to saving. To maintain any given level of income, planned investment must equal planned saving. Therefore, to provide an ever-growing level of income, an ever-growing percentage of the economy's production would have to be channeled into investment in order to absorb the ever-growing percentage of the economy's income that goes into saving. Whether business persons will in fact find it profitable to carry planned investment to the required level will, in the absence of government intervention, decide the crucial question of whether or not the economy can continue to grow and generate ever higher levels of income and output.

If, on the other hand, the basic consumption function is proportional, this problem does not appear. Growth in income alone will not affect the fraction of income that is saved. Hence, the fraction of the economy's production that must be devoted to investment remains the same. Assuming that government's relative role remains unchanged, the economy faces the much less serious problem of expanding investment proportionally with the growth of income and output rather than more than proportionally. Whereas the whole complex of conditions that determine the amount of investment expenditures planned by business persons might be sufficiently favorable to produce a growth of investment spending proportional with the growth of income, it is less likely that these conditions will be sufficiently favorable to produce a growth of investment spending that is more than proportional with the growth of income. The inability of the economy to absorb such saving would stop the growth of income and output dead in its tracks.

Although this particular problem no longer carries the same forbidding implications that it did forty years ago, we note it here to show that a practical question of great importance hinges in part on whether the basic consumption function is proportional or nonproportional. In the late thirties, before economists had the benefit of the data that later became available, the commonly held opinion followed Keynes in his belief that it was nonproportional. This contributed to another commonly held opinion, also suggested by Keynes, that the future would offer only more stagnation unless massive government deficits offset the volume of saving that was expected to grow more than proportionally with the growth of income. Of the several villains in this dismal picture, the belief in a nonproportional consumption function was one of no small importance.

The picture has changed over the years. There is now general agreement that the basic consumption function is proportional, which amounts to a rejection of the major tenet of the absolute income theory. If this indeed is the "true" nature of the long-run consumption function, we may expect consumption to continue to absorb a relatively constant proportion of an ever-growing national output of goods and services over the years ahead. We will not have to depend on uncertain special factors to shift a basically nonproportional long-run consumption function upward at the rate necessary to maintain consumption at a relatively constant proportion of output.

However, there is more to the explanation of consumption spending than the level of income. In the following chapter we will examine some other factors that influence consumption.

chapter 9
Other Factors Influencing Consumption Spending

The empirical evidence on the consumption–income relationship examined in the preceding chapter certainly suggests that disposable income is by far the most important factor influencing consumption spending. More specifically, it suggests that disposable income is of far greater importance in explaining consumption on a decade-to-decade basis than it is on a year-to-year basis, and of far greater importance on a year-to-year basis than on a quarter-to-quarter basis. In other words, the shorter the time period under consideration, the more clearly do factors other than the level of income leave their mark on the level of consumption spending.

Even if we exclude war periods in which the military absorbed a relatively large portion of the economy's total output, *quarter-to-quarter* changes in income are not infrequently accompanied by changes in consumption in the opposite direction. In such cases, the effect of the change in income is more than offset by nonincome factors influencing consumption in the opposite direction. However, apart from 1942 and 1945–47, *year-to-year* changes in income are accompanied by year-to-year changes in consumption in the same direction. Here the effect of the change in income more than offsets any nonincome factors influencing consumption in the opposite direction. Finally, not only have *decade-to-decade* changes in income since 1879 been accompanied by changes in consumption in the same direction, but, with the exception of the years of the Great Depression and World War II, the proportion of consumption to income in each decade has shown a remarkable stability. Thus, on a decade-to-decade basis, changes in income almost completely control changes in consumption.

What all this means is that changes in consumption spending, except perhaps from one decade to the next, cannot be explained solely on the basis of changes in disposable income. We cannot expect with any certainty that a change in income from one quarter to the next will be accompanied by a change in consumption in the same direction; the MPC is frequently negative. On the other hand, a change in income from one year to the next, if it is at all sizable, can with considerable certainty be expected to be accompanied by a change in consumption in the same direction. Even on a yearly basis, however, the change in consumption may be anywhere from a small to a large fraction of the change in income; the MPC is

positive but highly variable.[1] In short, for the changes we are most interested in—annual and quarterly—we are forced to recognize that income alone is not the complete explanation of the changes in consumption we observe in the real world. Other factors are clearly at work.

Because these other factors exert varying influence on consumption in each year, we find that the straight-line short-run consumption function for, say, 1967–76 does not pass through all the dots for these years in Figure 8-4.[2]

It is for this same reason that we find that the value for c, or 0.90, in the regression equation for this period, $C = 8.0 + 0.90Y$, does not tell us *for any particular pair of years* the amount of change in consumption that will accompany the indicated change in income from the first to the second of these two years. Instead, it tells us only the *average* relationship between changes in income and in consumption from year to year for the specific period of years in question.

From any perspective, the evidence clearly points to the presence of factors other than income that influence the amount that consumers spend in any year. While the income factor deserves major attention, the other factors are by no means of such little importance that they can be overlooked. In this chapter we will examine a few of them.

RATE OF INTEREST

While we may be reasonably certain that the rate of interest exerts some influence on the way in which any given level of aggregate disposable income is allocated between consumption and saving, we cannot say with equal certainty that a higher rate necessarily means that less of this income will be allocated to consumption and more to saving, or vice versa. The response to a change in the rate of interest may be an increase *or* a decrease in the total amount saved out of the existing level of disposable income.

By first examining the saving behavior of the individual, we may see why the response, in the aggregate, may go in either direction. Apart from that portion of an individual's saving that is prompted by a desire to leave an estate, the individual's current saving, which amounts to currently forgone consumption, may be viewed as deferred consumption. Since the typical individual has a positive time preference (meaning that he prefers a dollar of present consumption to a dollar of future consumption), he is willing to abstain from a dollar of current consumption only in exchange for something more than a dollar of future consumption. Assuming a stable price level over time, the rate of interest a person can receive on saving indicates the amount of future consumption he can secure in exchange. For example, at a rate of 5 percent, he can exchange one dollar of current consumption for approximately one and a half dollars of consumption eight years later. The individual aiming at intertemporal utility maximization will,

[1] As was shown in column (5) of Table 8-2, the MPC for year-to-year changes (omitting 1942–47) varied from a low of 0.340 in 1938 to a high of 12.000 in 1949. Large changes are found in adjoining years: from 0.395 in 1948 to 12.000 in 1949 and from 1.287 in 1969 to 0.461 in 1970.

[2] This requires qualification. Even if a straight line fitted to the plotted values for any period passed exactly through each of the values, we could not unqualifiedly conclude that only income determined consumption. There could be any number of other factors, some working toward higher and some toward lower consumption spending, whose net influence in each and every year canceled out. The improbability of such a result is, however, apparent. Beyond this, even in the absence of any other such factors, we still could not conclude that income alone determined consumption from the fact that the straight line passed through each of the values. As noted in the preceding chapter, the two variables are actually interdependent, and the single-equation approach here employed ignores this interdependency and simply assumes that consumption is the dependent variable and income the independent variable.

with a given interest rate, save the number of dollars out of his income such that the utility of the marginal dollar currently saved is just less than or just equal to the utility of the dollar sum to which that marginal dollar will have grown by a future date. He will, in other words, substitute future consumption for present consumption up to the point at which the marginal utility derived from the expenditure of the interest-augmented future dollar is just equal to or just greater than that derived from the expenditure of the current dollar.

Given the typical individual's schedule of preferences, it appears that the amount of saving out of his current income will therefore vary directly with the rate of interest. However, at the same time that a higher interest rate will tend to produce a substitution toward more future consumption and less present consumption, it also increases the individual's future income over what it would otherwise have been. Whatever the relationship between the individual's current income and expected future income had been, a higher interest rate raises his future income relative to his current income, which in turn may lead him to take part of this larger future income in the form of present consumption. Whether an individual with a given current income will save more on balance at a higher interest rate then depends on the relative strength of the *substitution effect*, which works toward more saving at higher interest rates, and the *income effect*, which works toward less saving at higher interest rates. For those lower-income individuals who will save only a relatively small part of their incomes even at high interest rates, the substitution effect will outweigh the income effect, and their saving will vary directly with the rate of interest. On the other hand, for individuals with large incomes who tend to save relatively large parts of these incomes, the opposite result may follow at high interest rates. The income effect may outweigh the substitution effect, with the result that higher interest rates will decrease the amount of current saving. For individuals in the former group, a supply curve relating the amount of saving to the rate of interest will slope upward to the right over any realistic range of interest rates. For individuals in the latter group, this supply curve may at some rate of interest display a backward bend, indicating that saving decreases with a rising interest rate for all rates above that at which this backward bend appears.[3]

There is a special case worth noting in which an individual's supply curve will be backward sloping not only at high rates of interest but at all rates of interest. This is the case when the saver's goal is simply the accumulation of a fixed dollar sum as of a particular future date and he saves whatever amount is needed to realize this goal. The higher the interest rate, the smaller then is the amount that he will find it necessary to save per time period in order to realize his goal.

The argument that a higher interest rate may cause some individuals to save more and others to save less receives its strongest support when limited to those individuals whose saving is some positive amount. The interest rate, however, may also affect the amount of saving of those individuals whose saving is some negative amount, and it may be expected for these individuals that a higher interest rate will in practically all cases discourage dissaving, which amounts to saying that it will encourage saving. Those individuals whose saving is a negative amount are financing current consumption in excess of current income by drawing down past savings or by borrowing. Because it is a rare individual who wants to borrow more at a higher rate of interest than at a lower rate, other things being equal, higher interest rates—which mean higher rates on installment credit purchases of cars and the like and on personal loans—will, if anything, discourage borrowing.[4] To the extent that the effect of higher rates is to make negative saving, or dissaving, smaller than it otherwise would be, it may be said that the

[3] For a detailed analysis in terms of indifference curves, see R. L. Crouch, *Macroeconomics*, Harcourt Brace Jovanovich, 1972, pp. 49–60, esp. 57–60.

[4] For the influence of interest rates on consumer demand, see M. J. Hamburger, "Interest Rates and the Demand for

amount of saving to this extent does vary directly with the rate of interest.

Although the supply curves of individuals who dissave will most likely vary directly with the interest rate for this reason, those of individuals who save may vary directly or inversely with the interest rate for reasons noted earlier. Therefore, the general shape of the aggregate supply curve, which is a summation of all these individual supply curves, can hardly be specified *a priori*. No simple, systematic relationship can be established between aggregate personal saving and the rate of interest. In view of this, many economists have taken an essentially agnostic position—they recognize that a change in the interest rate may change the amount of aggregate saving out of any given level of disposable income but that the direction of the change in saving, while most likely to be in the same direction as the change in the interest rate, may possibly be in the opposite direction.[5] This amounts to saying that interest-rate changes of the size found in actual experience do not, on balance, exert great influence on the amount of personal saving at any level of disposable income. There are forces working in both directions; these forces are in part offsetting and produce a net effect that is believed to be small.

The position that saving shows relatively little response to interest-rate changes has been widely held only since the thirties. It is quite different from that held by classical theory, in which the rate of interest was considered the most important factor influencing the amount of saving. Keynesian theory shifted the emphasis from the rate of interest to the level of income as the principal determinant of the amount the economy saves, regarding the former as of minor importance. As will be seen in Chapter 14, this is one of the essential differences between Keynesian and classical theory.

PRICE LEVEL AND PRICE EXPECTATIONS

Another factor that influences aggregate consumption expenditures is the price level of consumer goods and services. If it rises, will aggregate consumption expenditures increase or decrease? If it falls, will these expenditures increase or decrease?

To answer these questions, several distinctions must initially be made. First, we are concerned with aggregate consumption expenditures and not expenditures on any particular good or service or group of goods and services. A rise in the price of any single good or service for which there are close substitutes will lead to a transfer of expenditure to substitutes and a decrease in expenditures on the good that has risen in price. A decline in the price of such a good would, by the same reasoning, lead to a transfer of expenditures away from substitutes and increased expenditures on the good whose price has declined. However, for a rise or fall

Consumer Durable Goods," in *American Economic Review*, Dec. 1967, pp. 1131–53, and W. Weber, "The Effect of Interest Rates on Aggregate Consumption," in *American Economic Review*, Sept. 1970, pp. 591–600.

[5]This contention should not be confused with an altogether different one—namely, that saving will tend to be placed in those institutions or in those forms that pay the highest rate of interest consistent with savers' other objectives. This says that persons, *once having saved*, will tend to seek the highest rate of interest for their savings; it does not say that a higher rate of interest will necessarily induce more saving at any given level of disposable income. Thus, the intense competition, including "price (interest-rate) competition," among savings institutions does not necessarily influence the amount of income saved but may affect only its allocation among the competing institutions.

in the *price level* of consumer goods and services (e.g., the Consumer Price Index), there can be no goods or services that are substitutes or to which expenditures may be diverted.[6] The only substitution now possible is one of personal saving for consumption expenditures or consumption expenditures for personal saving. Consumers may, in other words, react to any change in the price level, rise or fall, by spending either more or less of their incomes for goods and services.

How consumers as a group may react depends in the main on a second distinction, that between a change in the price level accompanied by a proportional change in aggregate current-dollar disposable income and a change accompanied by a more or less than proportional change in aggregate current-dollar disposable income.[7] If current-dollar disposable income rises or falls proportionally with the rise or fall in the consumer price level, constant-dollar or real disposable income remains unchanged. If it rises more or less than proportionally, real disposable income, of course, increases or decreases, respectively.

Now, given a rise in the price level of consumer goods and services, will aggregate consumption expenditures rise or fall? If a rise (fall) in the consumer price level is offset by a proportional rise (fall) in aggregate current-dollar disposable income, consumers as a group are, in terms of real income, no better or worse off than they were before and presumably will hold real consumption expenditures and saving unchanged.[8] This is the conclusion we reach on the basis of the Keynesian argument that real consumption is a function of real income.[9]

"Money Illusion"

Although there is no change in real income, a change may nonetheless occur in real consumption if consumers are subject to what economists call "money illusion."[10] Suppose, for example, that during a given time period the consumer price level rises 10 percent and the current-dollar disposable income of each family rises 10 percent. Those families that recognize that their money income is unchanged in real terms suffer no money illusion, and, other things being equal, will probably maintain their

[6]Since a rise in the consumer price level does not involve proportional changes in the prices of all goods and services, some degree of substitution of goods that have risen relatively little or possibly even fallen in price for goods that have risen relatively sharply in price remains a possibility. However, it is also true that the cost to the average consumer of any likely assortment of goods and services, even after substitutions have been made wherever possible, will be higher after the rise in the price level than it was before. The greater the rise in the price level, the more certain this is to be true. This is all that is needed to qualify here as a rise in the price level as seen by the average consumer.

[7]There is no simple or unvarying relationship between a change in the consumer price level and a change in the level of current-dollar disposable income. It is possible for the two to change in opposite directions, although quarter-to-quarter data for recent years show, with few exceptions, changes in the same direction. Periods in which prices, not only of consumer goods and services but also of the other goods and services included in GNP, are rising are typically periods during which GNP in current dollars will be rising. Disposable income will rise along with GNP, but whether it rises more or less rapidly than GNP depends especially on how much of the enlarged gross income flow is diverted into tax receipts and retained corporate profits and how much the remaining gross income flow is supplemented by government transfer payments.

[8]Among other difficulties passed over here is the possible redistribution of current-dollar income between the fixed income group and other groups that may result from a changing price level. Incomes of some groups may rise or fall faster than prices, of others slower than prices, and of still others not at all. To the extent that the marginal propensity of these groups to consume differs, an impact on real consumption expenditures would be expected. Income redistribution as a specific factor influencing real consumption expenditures will be considered in a later section.

[9]Although attention is limited here to the effect of price-level changes, aggregate consumer spending as well as spending on individual goods and services may be affected by *relative* price changes quite apart from changes in the price level. (See G. Ackley and D. B. Suits, "Relative Price Changes and Aggregate Consumer Demand," in *American Economic Review*, Dec. 1950, pp. 785–804.)

[10]This term was coined by Irving Fisher, who used it to describe "a failure to perceive that the dollar or any other unit of money expands and shrinks in value." (See *The Money Illusion*, Adelphi, 1928, p. 4.)

consumption and saving unchanged in real terms. They will increase both spending and saving in current dollars by 10 percent. Other families may be subject to a money illusion in either of two ways. Some may see only that the price level has risen and somehow overlook the fact that their current-dollar disposable income has risen proportionally. They are actually no worse off, but they believe that they are and act accordingly. To the extent that families reduce the fraction of income saved in response to a decrease in real income, the families whose money illusion leads them to believe they have suffered a decrease in real income will reduce the fraction of income saved or increase the fraction consumed. This will, of course, involve an increase in real consumption, because their real income was actually unchanged. Other families may see only the rise in their current-dollar disposable income and overlook the proportional rise in the price level. These families will feel better off and, by the argument above, may increase the fraction of income saved or reduce the fraction consumed. This will then involve a decrease in real consumption, because their real income was actually unchanged.

If we assume that no widespread money illusion exists among consumers, we may expect no appreciable change in the fraction of aggregate current-dollar disposable income devoted to consumption expenditures as a result of a change in the consumer price level that is accompanied by a proportional change in current-dollar disposable income. In terms of the aggregate consumption function diagram, if we measure real income and real consumption on the two axes, the economy remains at the very same points on the income and consumption axes.

What about changes in the consumer price level that are not matched by compensating changes in current-dollar disposable income? Such changes mean changes in real disposable income, and a change in real disposable income should be expected to have a direct effect on real consumption expenditures. A rise in prices that leads to a fall in real disposable income should move consumers as a group back down the aggregate short-run consumption function. At this new point on the function, there is an absolute decrease in real consumption expenditures and an increase in the fraction of real disposable income devoted to consumption expenditures. Conversely, a fall in prices that leads to a rise in real disposable income moves consumers as a group up along the function with an absolute increase in real consumption expenditures and a decrease in the fraction of real disposable income devoted to consumption expenditures.

While these conclusions are valid, the changes in real consumption expenditures that result under these circumstances are not caused directly by the changes in the consumer price level but are brought about by the changes in real disposable income that accompany changes in the consumer price level. We already know from the empirical short-run consumption function that real consumption expenditures fluctuate with real income in this fashion. In short, we can say that a changing price level may affect real consumption expenditures to the extent that the changing price level is not offset by changes in current-dollar disposable income, provided that consumers are not subject to money illusion.

Price Expectations

In the preceding discussion, we were concerned with the relationship between *realized* changes in the consumer price level and *realized* changes in aggregate disposable income; we based our conclusion on the assumption that consumer spending behavior was determined *only* by the realized changes experienced by consumers in any time period. There are periods, however, in which the consumer price level is rising, current-dollar disposable income is rising to match prices, and real consumption expenditures are increasing. Note

that we are now suggesting an increase in real consumption expenditures with no change in real disposable income. This result frequently occurs when the consumer price level has been rising sharply for some time. It reflects the fact that consumers are devoting a larger fraction of real income to consumption expenditures "today" in *expectation* of still higher prices "tomorrow." Conversely, expectations of a lower price level in the future can lead to a postponement of real consumption expenditures currently, even though real disposable income has not fallen in the current period.

These changes in real consumption expenditures are the result not of the then current price level but of expectations as to what the future price level will be. Price-level expectations are but one of a number of types of expectation that may influence real consumption expenditures currently. For many families, an important factor may be expectations as to the future level of their income. Consumption expenditures are influenced to some extent also by the general expectations concerning the short-run outlook for expansion or contraction in business activity. Finally, any developments that give rise to expectations of large-scale war will produce expectations of shortages and higher prices. These expectations in turn will increase real consumption expenditures in the current period.

Consumer expectations with respect to all kinds of changes—economic, social, and political—can all have their effect on real consumption expenditures in any period. While noted here in connection with the price level, expectations in this more general sense may properly be considered as a separate factor influencing consumption. In this brief survey, however, space will not permit further discussion of expectations as a separate factor.[11]

DISTRIBUTION OF INCOME[12]

The level of aggregate disposable income is the most important factor that influences the level of aggregate consumption expenditures. However, with any given *level* of disposable income, the level of consumption expenditures resulting therefrom tends to be larger or smaller depending on the *distribution* of that income by income class. In general, the more equal that distribution is, the larger the fraction devoted to consumption tends to be. For example, an economy in which the 25 percent of the families with the highest income receive 30 percent of total income and the 25 percent of the families with the lowest incomes receive 20 percent of total income is more likely to show a larger frac-

[11] The major work on the influence of consumer expectations, attitudes and sentiment on consumer spending has been done at the Survey Research Center of the University of Michigan. This work began in 1946 and continues to date. See, for example, E. Mueller, "Effects of Consumer Attitudes on Purchases," in *American Economic Review*, Dec. 1957, pp. 946–65, and "Consumer Attitudes: Their Significance and Forecasting Value," in *The Quality and Significance of Anticipations Data*, National Bureau of Economic Research, Princeton Univ. Press, 1960, pp. 149–81. Other more recent studies include G. E. Angevine, "Forecasting Consumption with a Canadian Consumer Sentiment Measure," in *Canadian Journal of Economics*, May 1974, pp. 273–89, and G. Briscoe, "The Significance of Financial Expectations in Predicting Consumer Expenditures: A Quarterly Analysis," in *Applied Economics*, June 1976, pp. 99–119.

[12] Discussion here is limited to personal distribution or distribution by income group. For distribution by occupational group and functional share, see I. B. Kravis, "Relative Income Shares in Fact and Theory," in *American Economic Review*, Dec. 1959, pp. 917–49, and I. Friend and I. B. Kravis, "Entrepreneurial Income, Saving, and Investment," in *American Economic Review*, June 1957, pp. 269–301. At the same income levels, the marginal propensity of business owners and farmers to save is found to be greater than that of others. This suggests that shifts in the functional distribution of income over the business cycle may explain part of the short-run changes observed in consumption spending.

tion of that total income devoted to consumption than if the top 25 percent of the families received 45 percent of the total income and the bottom 25 percent of the families received only 5 percent of the total income. This is suggested by the fact that the fraction of disposable income allocated to consumption is higher at low levels of family income and lower at higher levels of family income. The family budget data of Table 8-1 gave us in Figure 8-1 an empirical family consumption function that showed this relationship.

While we may say that the more equal the distribution of income, the higher the fraction of income devoted to consumption tends to be, it does not necessarily follow from this that, other things being equal, any change in the distribution of a given level of income against higher-income families and in favor of lower-income families will mean an increase in the fraction of that level of aggregate disposable income that is allocated to consumption.[13] Although a redistribution would appear offhand to work in this direction, a number of qualifications that limit the quantitative importance of such changes and even question the direction of such changes must be noted:

1. The first is the fact that changes in the distribution of income in any short period, even a period as long as a decade, are moderate.[14] Since the existing occupational wage differences, distribution of property ownership, differences in productive ability, and the whole complex of institutional factors that determine the distribution of income show little change from year to year, changes in the distribution of income from year to year are correspondingly small and can have only a small effect on the fraction of aggregate disposable income devoted to consumption from year to year. Apparently the effect becomes significant only over a period of a decade or several decades. Of course, a deliberate public policy designed to reduce income inequality rapidly, regardless of its consequences, could alter this conclusion. Based on our past experience, however, even with a progressive tax structure and other governmental measures that have contributed noticeably to less inequality, the year-to-year, or even decade-to-decade, changes in income distribution have not been a very important factor influencing the amount of consumption expenditures at any given level of income.

2. Given any change in income distribution, a second qualification arises from the fact that the increase in the fraction of income devoted to consumption may be much less than is suggested by a comparison of the way that families at different income levels divide their total incomes between consumption and saving. Redistribution of a *given* aggregate income involves compensating changes in family income, additions to the incomes of some families and reductions in the incomes of others. Families do not ordinarily allocate an addition to or a reduction in their incomes between consumption and saving in the same way as they allocate their total incomes between consumption and saving. The *mpc* describes the allocation of additions to or reductions in the family's disposable income; the *apc* describes the allocation of the family's total disposable income. This distinction is important, because the differences in the

[13] Since the long-run trend in the distribution of income in the U.S. and other free economies has been toward less inequality, unless otherwise indicated all references to changes in distribution in the following pages will be to changes toward less inequality.

[14] The share of aggregate income received by the 20 percent of families with the lowest incomes changed from 4.5 percent in 1950 to 4.9 in 1960 and 5.5 in 1970. For the next higher quintile, the percentage was 12.0 in each of these three years. For the middle quintile, it was 17.4, 17.6, and 17.4 percent, respectively; for the next, 23.5, 23.6, and 23.5 percent, respectively; and for the top quintile, 42.6, 42.0, and 41.6 percent. (SOURCE: U.S. Department of Commerce, Bureau of the Census, *Current Population Reports*, Series P-60, No. 80, "Income in 1970 of Families and Persons in the United States," U.S. Government Printing Office, 1971, p. 28.) Although there were minor changes over the intervening years, note that the percentage was the same in 1950 and 1970 for all but the top and bottom quintiles. The highest quintile shows a decline of 1 percentage point and the lowest a corresponding rise of 1 percentage point over the 20-year period.

mpc at various family income levels are much smaller than the differences in the *apc* between these same family income levels.[15] If it is true that the immediate effect of any redistribution of income on aggregate consumption expenditures depends on the *mpc* rather than the *apc* at various family income levels, it follows that the prospective expansionary effect on consumption of a change in distribution may be overestimated if this effect is gauged from differences in the *apc* rather than from differences in the *mpc*.

In fact, it is possible that the *mpc* may be virtually the same at each level of family income. In such a case, even though the *apc* at higher incomes may be much below what it is at lower incomes, a change in distribution would have virtually no effect on consumption expenditures. According to some findings, something approaching this result may be found in actual U.S. experience.[16] Without pushing the argument to this extreme, however, the absence of pronounced differences in the *mpc* at different levels of family income has been adopted as one support for the argument that income redistribution, at least of the order realized in recent years, is a factor that does not significantly increase the fraction of disposable income allocated to consumption expenditures.

3. The effect on consumption to be expected from a given redistribution of income differs according to whether we follow the absolute, relative, or permanent income theory. As we will see, the relative income theory suggests that a decrease rather than an increase in consumption expenditures may be expected as a result of redistribution. The absolute income theory suggests just the opposite result, and the permanent income theory suggests that the amount of permanent consumption out of any given level of permanent income will be unaffected by a redistribution of that income.

1 Consider first the case in which spending and saving depend exclusively on absolute income. If a redistribution takes place with no change in the level of income, higher-income families will move down and lower-income families will move up the family consumption function in Figure 8-1. If each family regards its new position as permanent and immediately adjusts its spending to the amount that families at this income level customarily spend, we may expect an increase in the *apc* of the higher-income families after they move down and a decrease in the *apc* of the lower-income families after they move up. Since the *apc* of the lower-income families, though smaller after the redistribution than before, is still above the *apc* of the higher-income families, though this is larger after the redistribution than before, the net effect will be an increase in aggregate consumption expenditures at the given level of income.[17]

[15] Reference here is to the evidence supplied by budget studies that relate measured consumption to measured income as in Figure 8-1. There the moderate curvature of the empirical family consumption function indicates moderate differences in the *mpc* at different family income levels. If the function were a straight line, the *mpc* would be the same at all levels of family income. The *apc* at different levels of family income could be shown in Figure 8-1 by the slopes of a set of straight lines from the origin to various points on the consumption function. (Such a set of *apc* lines is given in Figure 8-2.) For any two levels of family income, a comparison of the difference between the slopes of the two lines from the origin with the difference in the slope of the consumption function itself at these two income levels will show the difference in the *apc* to be greater than the difference in the *mpc*.

[16] See H. Lubell, "Effects of Redistribution of Income on Consumers' Expenditures," in *American Economic Review*, March 1947, pp. 157–70, and "Correction," in *American Economic Review*, Dec. 1947, p. 930. See also M. Bronfenbrenner, T. Yamane, and C. H. Lee, "A Study in Redistribution and Consumption," in *Review of Economics and Statistics*, May 1955, pp. 149–59, M. Bronfenbrenner, *Income Distribution Theory*, Aldine, 1971, pp. 107–109, A. S. Blinder, "Distribution Effects and the Aggregate Consumption Function," in *Journal of Political Economy*, June 1975, pp. 447–75, and J. van Doorn, "Aggregate Consumption and the Distribution of Incomes," in *European Economic Review*, October 1975, pp. 417–23.

[17] For example, as a result of redistribution a \$21,000 family that had saved \$6,300 (*aps* = 0.30) and spent \$14,700 (*apc* = 0.70) becomes a \$20,000 family that now saves \$5,600 (*aps* = 0.28) and spends \$14,400 (*apc* = 0.72). A \$4,000 family that had saved nothing (*aps* = 0) and spent \$4,000 (*apc* = 1) now becomes a \$5,000 family that saves \$200 (*aps* = 0.04) and spends \$4,800 (*apc* = 0.96). The

2 Consider next the case in which relative income controls completely. The result is then quite different. With no change in aggregate disposable income, redistribution by definition means a change in relative incomes. All along the income scale, each family adjusts its spending to its new income level, not in accordance with the new absolute level of that income, but in relation to the spending patterns of the families whose consumption standards it emulates. As the income of the higher-income families is reduced through redistribution, the pressure on families just below to "keep up with the Joneses" is reduced downward in accordance with their reduced incomes. This process will exert its influence all the way down the income scale. The lowest-income families may still have no choice but to spend all their higher incomes, since these incomes are just sufficient to cover basic necessities. Other families, however, may now find it possible to save a part of their income that was previously used to purchase the fancier automobile, the trip abroad, the backyard swimming pool, and other such goods and services that were largely emulative purchases. A redistribution, instead of leading to increased consumption, may actually decrease consumption, because the lessening of inequality wipes out some of society's emulative spending.

3 Finally, the case in which permanent consumption is related to permanent income in the way indicated by the permanent income theory, gives yet a different result. Recall that this theory holds that the ratio of permanent consumption to permanent income is the same at all levels of family income and that there is zero correlation between transitory consumption and transitory income. A redistribution toward greater equality will provide increments of income to lower-income families and the opposite for higher-income families. These additions to and subtractions from income may at first be regarded by these families as positive and negative transitory-income components. In this case, the permanent income theory holds that neither the lower-income families nor the higher-income families will change the amount of their consumption, so that the redistribution is without effect on consumption. A program of redistribution that is maintained year after year will eventually cause the families gaining or losing income to regard these as permanent changes in income and will thus raise the permanent incomes of lower-income families above what they otherwise would have been. However, the aggregate of permanent consumption is unaffected to the extent that the ratio of permanent consumption to permanent income is, as the theory holds, the same for all otherwise similar families, regardless of the level of their permanent income. The result therefore appears to be the same in the short period, when the redistribution may be viewed by those who gain or lose from it as transitory changes in income, and in the longer period, when the redistribution is viewed by those who gain or lose from it as permanent changes in income.

Since the several theories suggest quite different results, one cannot choose one result over another without first choosing one theory over another. And the relative merits of the competing theories continue to be sharply debated. Few economists so far have accepted the permanent income theory in its extreme form, and therefore few economists would deny that redistribution has some effect on consumption. However, the importance of this effect is a matter on which there is little disagreement. In contrast to the belief generally held in the early years following the appearance of Keynes's *General Theory*, economists no longer believe that raising the level of aggregate consumption spending by redistribution is the simple matter they once thought it to be.

increase in saving of the low-income family will not be as great as the decrease in saving of the high-income family; the decrease in consumption of the high-income family will not be as great as the increase in consumption of the low-income family. Thus, the two families that showed total consumption of $18,700 and total saving of $6,300 before the redistribution show total consumption of $19,200 and total saving of $5,800 after the redistribution, although the combined income of the two families is $25,000 in both cases.

FINANCIAL ASSETS

Some economists have attached considerable importance to the volume of financial assets accumulated by consumers as a factor influencing consumption expenditures. It seems reasonable to expect that, on the average, the family with a larger accumulation of cash, demand deposits, savings deposits, stocks, bonds, and other kinds of financial assets will spend more than the family with a smaller accumulation of financial assets, other things (including the disposable income of each family) being equal.[18] The rationale here is that a family's urge to add still more to its holdings of financial assets diminishes as its holdings of these assets increase. If most families reacted this way, as the total holdings of financial assets of all families increased over the years, a larger fraction of aggregate disposable income would be devoted to consumption and smaller fraction to saving or to the accumulation of still more financial assets. In terms of the aggregate consumption function, this suggests that the entire curve would shift upward with the growth in the volume of financial assets, thereby producing a higher APC at each possible level of aggregate disposable income.

As a case in point, the upward shift in the aggregate consumption function in the first of the postwar periods shown in Figure 8-4 may be attributed to the huge amount of savings bonds, bank deposits, and other liquid assets accumulated by consumers during World War II. Recall that the upward shift in consumption was not limited to the first few postwar years, which could be readily attributed to spending to satisfy deferred demand. Instead, this shift characterizes later years as well. It may be argued that during the war years families built up a volume of liquid assets that they regarded as adequate or more than adequate to meet any future emergencies. With this accumulation of cash equivalents to fall back on in case of need, families felt free in the postwar period to devote a larger fraction of their current disposable income to consumption. In contrast, a smaller fraction of disposable income had been devoted to consumption in the prewar period, for then only a small minority of the nation's families had an appreciable amount of such assets.

At first glance this explanation suggests a clear-cut connection between the volume of financial assets held by consumers and the fraction of their disposable income devoted to consumption. As before, however, qualifications appear as soon as we examine this theory a little more closely.

For one thing, the size of the consumption-inducing effect depends on the ownership distribution of these financial assets by income class. 1 An increase in the volume of these assets, if concentrated in the hands of upper-income families, may give little or no stimulus to consumption. Upper-income families save a large fraction of their incomes in any event, and it is doubtful whether the further growth in their holdings of financial assets would cause the fraction of their income that is saved to decrease.

A related qualification, which applies regardless of ownership distribution, is the possibility 2 that for many families a taste of financial assets simply whets the appetite for more. Even fam-

[18]As a factor influencing consumption expenditures, economists usually count only "liquid assets," which exclude any financial assets, such as conventional mortgages, that cannot be quickly converted into cash at fixed prices. The argument is that it is liquid assets more than illiquid assets that influence consumption because the possession of the former allows a consumer to dissave very readily. Some economists, however, include all financial assets and also make allowance for financial liabilities of consumers, the major part of which is debt outstanding as the result of purchases of houses and durable consumer goods. Beyond this, one need not limit attention to financial assets but may include real assets like housing and personal property and may allow for all liabilities to derive a net worth figure as in a consumer's balance sheet. However, the brief discussion here will not go beyond consideration of financial assets alone.

ilies of moderate income may react this way. Their original preference for consumption over saving may, after the first taste of financial assets, change to a preference toward saving as the means of accelerating the accumulation of financial assets.[19] Although this may be far from common for many families, particularly for low-income families, it is still a factor to be considered. This in itself is sufficient to raise some doubt as to the direct connection posited above between the volume of financial assets held by consumers and the level of consumer expenditures at any given level of income.

Finally, the mere dollar volume of financial assets may suggest incorrect relationships; only the real value of the dollar volume held at any time is pertinent. A changing price level will cause the real value of a given volume of financial assets to rise or fall. Except for holdings of stocks, shares in mutual funds, or other such assets whose current-dollar valuation tends to rise and fall with the price level, all other financial assets are fixed-dollar assets. A family with $10,000 in fixed-dollar assets will, barring money illusion, recognize a rise in the price level as a decrease in the real value of its fixed-dollar assets and a fall in the price level as an increase in the real value of these assets. For example, a doubling in the price level will be seen to cut in half the purchasing power of the $10,000. If a rise in the price level is matched by a proportional rise in the family's current-dollar disposable income so that its real income remains unchanged, it may nevertheless cause a reduction in its real consumption expenditures due to the depressing effect of the price-level change on the real value of its holdings of fixed-dollar assets. Conversely, if a fall in the price level is matched by a proportional fall in the family's current-dollar disposable income, it may still increase its real consumption expenditures due to the expansionary effect exerted by the rise in the real value of its holdings of fixed-dollar assets.

The consumption-stimulating effect of an increase in the real value of financial assets brought about by a fall in the price level has been named the *Pigou effect* after A. C. Pigou, the eminent British economist who first clearly set forth the relationships between consumption, financial assets, and the price level.[20] We will examine the Pigou effect in more detail in Chapter 17 and find that some writers have assigned it a highly important theoretical role that goes beyond that of a minor influence on consumption spending.

Of a related nature but still distinct from the Pigou effect is the consumption-stimulating effect of an increase in the real value of financial assets brought about by a fall in interest rates. Outstanding debt obligations that promise to pay a fixed number of dollars of interest each year and the principal amount at maturity will vary inversely in value with market interest rates.[21] For example, a family that today holds a high-grade, 8 percent corporate bond with a $10,000 face value and maturity in the year 2005 may expect to receive $800 interest each year and the $10,000 principal in the year 2005. If market interest rates on bonds of this kind were to rise substantially between now and, say, 1985, the family would experience a sizable decrease in the price at which it could sell that bond and thus a sizable decrease in the real value of that bond as of that date. In the opposite case of a substantial decline in market interest rates on this kind of bond as of 1985, the family would experience a sizable increase in

[19] Note that a revision of preferences toward saving and away from consumption is also one toward income and away from consumption. Less consumption and more saving, with the saving devoted to the acquisition of income-producing financial assets, is a means of raising the level of family income and making possible still more saving.

[20] A. C. Pigou, "The Classical Stationary State," in *Economic Journal*, Dec. 1943. See also his "Economic Progress in a Stable Environment," in *Economica*, Aug. 1947, reprinted in F. A. Lutz and L. W. Mints, eds., *Readings in Monetary Theory*, Irwin, 1951, pp. 241–51.

[21] See pp. 261–63 for a brief explanation of this inverse relationship between interest rates and the value of debt obligations.

the real value of that bond as of that date.[22] In the first case, the decrease in the real value of its financial assets could cause the family to reduce the amount of consumption out of an unchanged real income, and in the other case the opposite, of course, would be true. This interest-induced effect on consumption is seen to parallel the price-induced effect, or Pigou effect, just described.

Depending on the quantity of financial assets held and the magnitude of changes in interest rates, the real value of these assets may change considerably over time. Therefore, potentially at least, changes in interest rates could have a significant impact on consumption spending. This is an interesting possibility, because it means that the interest rate is at least potentially important as an indirect influence on consumption spending through its effect on the value of assets, whereas it does not appear to be at all important as a direct influence. In other words, although people may not be directly induced to save a larger portion of their income by the fact that a higher rate of interest means a greater return on their saving, they may be indirectly induced to do so by the fact that a higher rate of interest means a decrease in the value of some of the assets they already hold.

Our immediate interest in this section is with the influence of the amount of financial assets held by consumers on the level of their spending, but, as we have now seen, any consideration of this factor is closely related to two other factors discussed earlier, the price level and the rate of interest. The real value of the financial assets held by consumers changes not only because consumers devote part of their saving in each period to the acquisition of more such assets but also as a result of revaluations due to changes in the price level and the interest rate. But whatever the cause of any particular change in the real value of these assets, the relevant question here is whether such a change produces a change in consumption spending. There is little doubt that there is a direct relationship here, but the quantitative strength of the relationship is another matter.[23] The evidence is not clear, but it would appear that these assets are not ordinarily a major influence on consumption spending.[24] The holding of such assets is, nonetheless, one of the major *nonincome* factors that influence consumption spending and as such warrants our attention in this chapter.

OTHER FACTORS, IN BRIEF

There are a number of other nonincome factors that affect consumption expenditures; a few of them will be mentioned here.

In any time period, relatively easy consumer credit terms can stimulate consumer expenditures for durable goods, and relatively tight

[22]Although our interest here is in financial assets, the relationship in question applies in general to all assets that produce fixed income streams. A rise or fall in market interest rates means a rise or fall in the rate at which the income stream from any such asset will be capitalized. For example, the owners of an office building that produces a given income flow per time period (under long-term leases) would see the price at which they could sell this building decline in the face of rising interest rates, other things being equal.

[23]For a quantitative estimate of the influence that consumers' liquid asset holdings have on consumption expenditures, see A. Zellner, "The Short-Run Consumption Function," in *Econometrica*, Oct. 1957, pp. 552–67, and "Further Analysis of the Short-Run Consumption Function with Emphasis on Liquid Assets," in *Econometrica*, July 1965, pp. 571–81. See also K. Marwah, "Measuring the Role of Liquid Assets in Consumption: A Cross Country View of the World Economic Periphery," in *Journal of Development Studies*, April–July 1974, pp. 332–46. A list of consumption functions with liquid asset terms included is provided in D. Patinkin, *Money, Interest and Prices*, 2nd ed., Harper & Row, 1965, Appendix M.

[24]For an introductory statement on the general concept of the price-induced and interest-induced wealth effects

terms can do the opposite.[25] The explosion in consumer purchases of automobiles far back in 1955 is credited in large part to the unusually easy credit terms offered in that year. When we recall that an increase in expenditures for durables tends to be largely at the expense of a decrease in saving, we can understand how easy credit, which stimulates purchases of durables, can increase the fraction of aggregate disposable income that is spent in any time period.[26]

The difference between the rates of growth of aggregate disposable income and population over a period of years is another factor that influences the fraction of aggregate disposable income devoted to consumption. If the nation's population is growing faster than aggregate disposable income, disposable income per capita is decreasing, and this in turn tends to raise the fraction of aggregate disposable income that is consumed. Changes in certain characteristics of the population, such as its age distribution, will also affect the fraction of aggregate income spent. For example, an increase in the percentage of the population that is made up of persons past working age or under working age will tend to increase the percentage of disposable income spent, since the average propensity to consume in these age groups is higher than the average for all age groups combined.

Deferred demand, especially on the scale displayed in the years just following World War II, tends to push consumption expenditures above the level that would be expected on the basis of disposable income alone. But it is apparent that this factor has little application in normal peacetime periods.

Over recent decades, institutional changes have occurred that also affect the spending–saving patterns of consumers. For many families, saving has become largely automatic with the popularization of long-term saving commitments through life insurance, private pension funds, and mortgage loans on homes, the repayment of which is amortized over a long period of years. Once families are committed, the amount of such quasi-compulsory saving is largely immune to changes in income. Since income does change, the stability of this portion of personal saving is another factor influencing the fraction of income that is devoted to consumption. For the portion of saving that remains a matter of current decision by the family each week or month, the diversity of forms in which personal saving may be placed is now so great that a saver may find an outlet for his saving that is virtually tailor-made to his requirements. To some degree this also affects the spending–saving patterns of consumers.

CONSUMPTION DEMAND—A CONCLUDING NOTE

In attempting to explain what determines the level of consumption spending, one is tempted, as were many economists in the early years following the appearance of Keynes's *General*

and an argument supporting the importance of assets or wealth on consumption spending, see B. P. Pesek and T. R. Saving, *Money, Wealth, and Economic Theory*, Macmillan, 1967, pp. 11–21. See also M. J. Bailey, *National Income and the Price Level*, McGraw-Hill, 2nd ed., 1971, Ch. 6.

[25]Easier consumer credit is not the same as a lower rate of interest, although the latter may be one characteristic of easier consumer credit. Consumer installment credit may be said to be "easier," with no reduction in rates charged, if there is a lengthening of the repayment period on installment purchase contracts or if there is a relaxation of credit standards that makes credit available to applicants previously rejected.

[26]See P. W. McCracken, J. C. T. Mao, and C. Fricke, *Consumer Instalment Credit and Public Policy*, Bureau of Business Research, Univ. of Michigan, 1965. An appendix provides a case study of the role of installment credit in the 1955 auto sales year. See also D. B. Eastwood and R. C. Anderson, "Consumer Credit and Consumer Demand for Automobiles," in *Journal of Finance*, March 1976, pp. 113–23.

Theory, to begin and end the explanation with the level of disposable income as *the* determinant. The time series data that first became available for the U.S. economy during World War II gave empirical support to disposable income as the major factor in any explanation, but at the same time these data showed disposable income to be something less than a total explanation. Many other factors were found to influence consumption spending, particularly in the short run, and these dispelled any notion of a stable relationship between changes in consumption and changes in disposable income on a quarter-to-quarter or even a year-to-year basis. However, despite the influence of these other factors, the data still clearly suggested that the level of disposable income dwarfs in importance any nonincome factors and, except for short-term changes in income and consumption, is more important than all these nonincome factors combined.

This dominance of the income factor permits economists to build meaningful theoretical models in which consumption is made a function of income alone. This does not mean that consumption equations that allow for variables other than income are not superior as explanations of consumption spending. Quite the contrary. Nor does it mean that an equation that includes only income as an explanation of consumption need be as simple as $C = C_a + cY$. At the minimum, it must be recognized that, to the extent that consumption in any time period does depend on income, it depends on income of a number of other time periods as well as that of the current time period. The equations that express the relative, permanent, and life-cycle theories of consumption all take this into account and in so doing become quite complex.

For our purposes in the chapters ahead, we will limit ourselves to the equation in the form $C = C_a + cY$, in which C and Y alone are variables and C_a and c are constants. Therefore, any change in consumption spending from one period to the next, ΔC, will be determined solely by the change in disposable income from one period to the next, ΔY. In other words, holding C_a and c unchanged means that no further attention will be paid to changes in the various nonincome factors that actually affect the height of the function, C_a, and its slope, c. By assuming that these nonincome factors remain unchanged, we have an unchanging consumption function whose height and slope are given, so that consumption can change only as a result of a change in disposable income. Thus, consumption is assumed to be entirely passive, responding to changes in income but not initiating them, as would be the case if the function were assumed to shift upward or downward with changes in nonincome factors.

One purpose of this chapter has been to provide some understanding of what the nonincome variables are, what their importance is, and how each one may influence consumption spending. Such an understanding puts into perspective the simplified relationship in which consumption is assumed to be a function solely of the income variable. In an introductory treatment of macroeconomic theory, we must focus exclusively on the income variable to simplify what would otherwise be an exceedingly complex relationship. However, in the chapters ahead, it is important to keep in mind that, although income is certainly the all-important variable influencing consumption and the only variable that will appear explicitly in the models, consumption in practice depends also on a large number of nonincome variables.

chapter

10 Capital and Investment

In a two-sector economy consisting only of consumers and businesses, aggregate spending for goods and services equals the sum of consumer spending and business spending, or the sum of personal consumption expenditures and private domestic investment expenditures. A change in aggregate spending may be initiated by a change in either consumption expenditures or investment expenditures, but it is more commonly the latter that appears to be the source of the changes in spending that set into motion sizable movements in the income level. With changes in investment playing this strategic role, it follows that an explanation of the changes that occur in income requires that we first find an explanation for the frequent and sometimes drastic changes that take place in investment.

In seeking the explanation, economists have not been able to identify a single dominant variable as they have for consumption. Although disposable personal income, as we have seen, is by no means the total explanation of consumption, there is little disagreement as to its primacy. Within the area of investment, there are those economists whose research leads them to the conclusion that primacy must be assigned to profits and those whose research leads them to the quite different conclusion that primacy must be assigned to the acceleration principle or in some way to the matter of capacity utilization. None of these economists, of course, argues that either of these factors taken by itself provides a complete theory of investment, but the proponents of each do argue that one alone is sufficient on which to build an adequate theory of investment. In contrast to these two groups are those economists whose work leads them to favor a compromise approach that incorporates both profits and the acceleration principle as necessary elements. In their judgment, an adequate explanation of investment requires both of these elements and perhaps others as well—for example, the rate of interest and other financial considerations.

With no one factor unanimously recognized as dominant or by itself adequate as an explanation, we cannot approach the study of investment spending by focusing almost exclusively on one factor as we did in the study of consumption spending. We will find it advisable as a minimum to give attention to the flow of profits,

the acceleration principle, and interest rates and the role of finance. The task of explaining how these factors affect investment expenditures is undertaken in the next two chapters. However, before getting into that we must lay some groundwork in the present chapter.

Investment is the acquisition of capital assets, and the first task of this chapter is to make clear the basic relationship between the flow called investment and the stock called capital. The firm's acquisition of a capital asset, whether a small piece of machinery or a whole new plant and its equipment, is based on considerations quite different from those that underlie an individual's purchase of consumer goods of any kind. The firm's decision to purchase or not purchase in the case of capital goods requires a calculation of the rate of return expected from the asset over its life. Our second task in this chapter is to see how the firm calculates the rate of return on a capital asset and how this rate relative to the rate of interest underlies the decision to invest. In the last part of the chapter we will look into the process by which a discrepancy between the actual amount of capital held by business and the amount it desires to have is made up step by step over time through net investment spending.

THE MEANING OF CAPITAL AND INVESTMENT

Investment, a word with many meanings in popular usage, has only one meaning in national income analysis—the value of that part of the economy's output for any time period that takes the form of new structures, new producers' durable equipment, and change in inventories. In practice, apart from the change in inventories, the value of this output is measured by the amount of expenditure on these items. Investment so defined can be viewed, we recall, in either gross or net terms. If we deduct from gross investment expenditures an allowance for the amount of plant and durable equipment used up in turning out the period's output, we have net investment.

In any attempt to explain aggregate investment expenditures in any time period, one difficulty is the fact that different factors determine the different types of investment expenditure. No single investment theory can reasonably apply to all forms of investment expenditure. The amount of expenditures by persons for owner-occupied housing is not dominated by profit considerations as is the amount of expenditures by business persons for commercial and industrial structures and for durable equipment. Similarly, the amount of expenditures by business persons for new plant and equipment is affected by factors different from those that determine the amount they spend for additions to inventories, even though it is the expectation of profit that dominates in both cases. Although some attention will be paid to other types of investment in Chapter 12, in this chapter investment will refer specifically to business expenditures for plant and equipment. In recent years these components have made up roughly two-thirds of gross private domestic investment.

Investment is a flow variable whose counterpart stock variable is capital. Capital, another word with many meanings, should here be understood to mean only the accumulated stock of plant and equipment held by business. If, for the economy as a whole, gross investment in any period equals the amount of capital used up during that period, there is neither net investment nor disinvestment and so no change in the stock of capital. If gross investment exceeds replacement requirements, the difference equals positive net investment, which represents an increase in the stock of capital. If gross investment is less than replacement require-

ments, the difference is negative net investment, or disinvestment, which represents a decrease in the stock of capital.

Therefore, by definition, net investment is an addition to the stock of capital. Other things being equal, an addition to the stock of capital means an increase in the productive capacity of the economy. This must be the result when a larger physical stock of capital is available for use with an existing labor force, natural resources, and technology. In the same way, an increase in productive capacity must result when there is an increase in the labor force with no change in the stock of capital, natural resources, or technology or when there is an improvement in technology with no change in the stock of capital, labor force, or natural resources. As suggested by the "law of diminishing returns," the only plausible exception to this rule is the case in which the variable factor is so plentiful relative to the fixed factors that its marginal productivity falls to zero. This situation is very unusual.

The stock of capital for the economy as a whole tends to grow over time, and over time the labor force, at least so far, also has tended to grow and technology to improve. In practice, other things do not remain equal as the capital stock grows. However, for analytical purposes, it is necessary to isolate the growth in the capital stock from the growth in other factors in order to specify the increase in potential output associated with the growth in the capital stock. To assume an unchanged technology or a given "state of the arts" presents no serious conceptual difficulty; we are simply to understand that the same methods of production are employed. Net investment and the resultant growth in the stock of capital then means that more of the existing kinds of capital goods are employed in the same way as existing capital goods. To assume an unchanged labor force does, however, present some conceptual difficulty. While the stock of capital might conceivably grow virtually without limit even in the absence of technological change, it is more difficult to imagine unlimited growth in the stock of capital of this kind without a growth in the labor force. More workers would be needed to utilize an enlarged stock of capital that involves no labor-saving changes in technology.

If the stock of capital, the state of technology, and the labor force are all treated as variables, this conceptual difficulty disappears and others appear. With all treated as variables, it is possible to have technological improvements that lead to a growth in the stock of capital without any corresponding increase in the amount of labor required for the efficient use of the now larger stock of capital. With the state of technology as a variable, we may also have a growth in the stock of capital that is intended not as a means of increasing capacity but rather as a means of reducing the cost of producing the level of output attained with existing capacity. "Modernization" of a firm's plant and equipment through net investment expenditures may amount to the adoption of technologically superior and therefore lower cost-per-unit methods of producing an unchanged level of output. Such cost-cutting net investment may be distinguished conceptually from net investment that is designed simply to expand the productive capacity of the firm's plant and equipment. However, for the economy as a whole, whether net investment is undertaken for "modernization" or for capacity expansion, all net investment increases the productive capacity of the economy as a whole. If a plant is "modernized" in order to permit a given level of output to be produced with less labor, the labor so released is available to enlarge production elsewhere in the economy and thus enlarges the economy's productive capacity.

To avoid the complications inherent in simultaneously treating all these factors as variables, we will assume for the present that the state of technology, the size of the labor force, and the amount of the economy's natural resources are all given. As we develop the theory of net investment spending for the economy as a whole, we may simply identify the resultant expansion in the capital stock as an expansion in the productive capacity of the economy.

THE DECISION TO INVEST

Since most capital goods remain useful for many years, one can learn only after several years have passed whether the investment in these long-lived goods will turn out to be profitable or unprofitable. However, if a particular investment expenditure is made, for replacement as well as for expansion, it generally means that the business person who invests has estimated that the investment will be profitable.[1]

Three Elements Involved

The business person's estimate of the profit or loss that will accrue from any particular investment is based on the relationships among three elements: the expected income flow from the capital good in question, the purchase price of that good, and the market rate of interest. Because a forecast of what lies in the uncertain future is unavoidable, the crucial factor in the business person's evaluation of the prospective profitability of any investment expenditure is his estimate of the income flow the capital good will yield over its life. This uncertainty arises not only from uncertainty as to the amount of income the good will generate each year of its life but also from the number of years in that life. The good may turn out to be less durable than was originally anticipated, or it may become obsolete due to technological changes well before it is physically worn out. Beyond this, in some cases even the purchase price is uncertain. If instead of a single capital good like a piece of machinery, we think of the construction and equipping of a complete new plant, its full cost to the firm at the time of its completion several years later may well turn out to be substantially different from what was anticipated at the time the decision to go ahead with the project was made.

To trace through the basic relationships among the three elements involved in the investment decision, let us for the moment ignore the matter of uncertainty. Suppose, then, that management estimates that a particular machine whose purchase it is considering can remain in use for a five-year period, at which time it will have only a negligible scrap value. Management will estimate its physical productivity—the increase in the number of units of final output made possible in each of the five years because of the addition of this machine to its stock of capital. Then to obtain the figure for the gross income flow expected from the machine each year, the estimated marginal physical productivity for each year is multiplied by the estimated price per unit (net of indirect taxes) at which each year's additional output can be sold or, more exactly, is multiplied by the marginal revenue per unit. However, in producing and selling this additional output, extra raw materials, power, advertising, and labor will probably be required.[2] When the total cost of these extra inputs for each year is subtracted from the estimated gross income flow for each year, the remainder is the estimated net income pro-

[1] We make the standard assumption that the firm's objective is "profit maximization." Although it is well known that firms are influenced by other objectives, for simplicity "profit maximization" will be assumed here to be the exclusive objective. For an introduction to the question of the validity of this assumption, see R. M. Scherer, *Industrial Market Structure and Economic Performance*, Rand McNally, 1970, pp. 27–36. A list of some of the leading contributions to the literature in this area is found on p. 27 of this book. See especially F. Machlup, "Theories of the Firm: Marginalist, Behavioral, Managerial," in *American Economic Review*, March 1966, pp. 1–33.

[2] Although we have assumed that the labor force, natural resources, and technology are all unchanged, the individual firm can always get more labor and raw materials by bidding them away from other firms. This is patently impossible for all firms together, but all firms may still redeploy the available labor force and the supply of raw materials so as to make optimum use of them in combination with a growing stock of capital. Of course, as all firms taken together increase the stock of capital, the additional output per additional unit of capital must eventually decrease, if the quantity of labor and other factors remains unchanged. Diminishing returns enter the picture.

duced by the machine in each year. In computing this figure, all costs that will be incurred in using the machine and in producing and selling its output are deducted, with two notable exceptions: (1) the annual depreciation expense, which on a straight-line basis would equal one-fifth of the purchase price of the machine, and (2) the annual interest cost on the amount of funds tied up in the machine, a cost equal to the undepreciated portion of the purchase price times the market rate of interest. Thus, if in the first year one-fifth of the original outlay is covered by setting aside in a depreciation reserve one-fifth of the purchase price of the capital good, the interest cost in the second year will equal the market rate of interest times four-fifths of the purchase price. The estimated net income figures for the five years make up a series of figures that may be designated as R_1, R_2, R_3, R_4, and R_5.[3]

Suppose that the sum of R_1, R_2, R_3, R_4, and R_5 exceeds the purchase price of the capital good. Can this excess over the amount required to replace the machine be taken as the dollar amount of estimated profit from the machine over its five-year life? Can we divide this figure by 5 and call this amount the average profit per year? Can we divide this average amount by the purchase price of the machine and call this the rate of return on the investment? The answer to all these questions is no. Not all the excess of the income flow over the purchase price is profit. We made no allowance for the fact that the income flow from the capital good will trickle in over the next five years, while the outlay for the capital good is made in one lump sum at the time of purchase. To disregard the time difference between income and outlay would be to equate today's dollar with tomorrow's; as long as a positive rate of interest can be earned by any lender, a dollar to be received tomorrow will necessarily be worth less than one received today. To compare properly the number of dollars paid out today in purchasing the capital good with the number of dollars of income that the good will earn over its life, we must compute the *present value* of those future dollars. This requires that we *discount* each of those future dollars for the number of years it is removed from the present; the present value depends on the rate at which the future dollars are to be discounted. Discounted at 4 percent, the sum of $100 to be received five years from today has a present value of $82.19; at 5 percent the same $100 has the smaller present value of $78.35; and at 6 percent the still smaller present value of $74.73. Part A of Figure 10-1 shows this graphically. The higher the rate that is used, the lower is the present value. In the interesting but nonexistent case of a zero rate, the present value is seen to be identical with the future value.

The present value of any future dollar also varies inversely with the time period involved. Assuming a 6 percent rate, the present value of $100 to be received five years from today is $74.73, but the present value of the same number of dollars to be received one year from today is $94.34. Part B of Figure 10-1 shows the present value of $100 to be received 1, 2, 3, 4, and 5 years in the future, in each case discounting at 6 percent.

This discounting process—the process by which a future sum shrinks as it is translated into present value—is simply the reverse of the more familiar process by which a sum grows as it is carried into the future. If the present value of $100 to be received five years from today is $82.19 when it is discounted at 4 percent, $82.19 put out today at 4 percent interest will grow to $100 at the end of five years. This is shown by the uppermost curve of Part A, in which we visualize the growth over time of $82.19 today to $100 five years from today. Similarly, if the present value of $100 to be received four years from today is $79.21 when discounted at 6 percent, then $79.21 put out today at 6 percent interest will grow to $100 at the end of four years. This is shown by the curve for the four-year period in Part B.

[3]Of course, if we choose to express net income after income taxes, a deduction for this would have to be made in computing the R values.

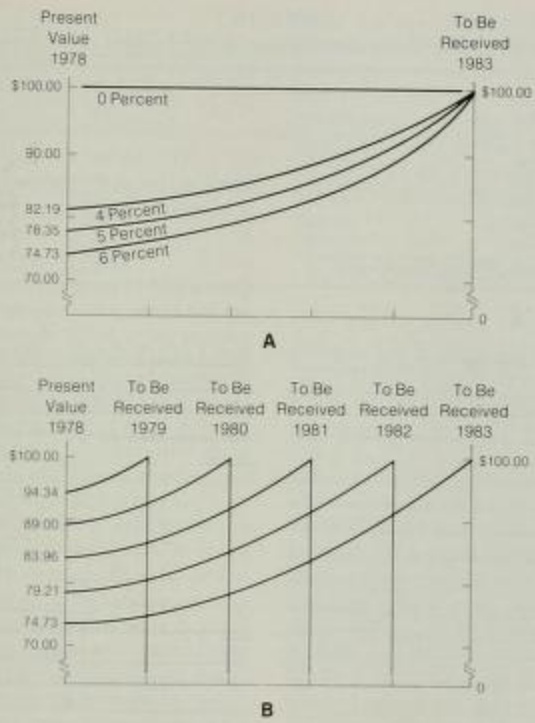


FIGURE 10-1
Present value of $100 for selected discount rates and time periods

Let us assign specific dollar values to R_1, R_2, R_3, R_4, and R_5 and determine the present values of these amounts at 4, 5, and 6 percent discount rates. Suppose that for each of the five years the estimated income from the capital good is $100 after deduction of all expenses other than interest and depreciation. (To simplify, assume also that each year's income appears in one lump sum at the end of that year.) The results for each of the three discount rates are given in Table 10-1. Note that the series of five figures in the final column corresponds with the five figures at the left side of Part B of Figure 10-1.

The total given for the present values at the bottom of each column is the amount which, if put out today at the rate at the head of each column, would generate a stream of income of $100 at the end of each of the next five years and then nothing.[4] That is, this income stream

[4]Thus, if one invests $445.18 at 4 percent for one year, he earns $17.81 in interest for that year, making his investment at the end of the first year $462.99. Withdrawing $100 from investment at the end of the first year, he invests $362.99 for the second year, on which he receives $14.52 of interest, making his investment at the end of the second year $377.51. Withdrawing another $100 from investment, he invests $277.51 for the third year, on which he receives $11.10 in interest, making his investment at the end of the

TABLE 10-1
Present value of an income stream at various rates of discount

ESTIMATED INCOME	4%	DISCOUNTED AT 5%	6%
R_1 $100 at end of 1st year	$ 96.15	$ 95.24	$ 94.34
R_2 $100 at end of 2nd year	92.46	90.70	89.00
R_3 $100 at end of 3rd year	88.90	86.38	83.96
R_4 $100 at end of 4th year	85.48	82.27	79.21
R_5 $100 at end of 5th year	82.19	78.35	74.73
Total of Present Values	$445.18	$432.94	$421.24

provides a yield of 4 percent if the machine can be purchased for $445.18, a yield of 5 percent if it can be purchased for $432.94, and a yield of 6 percent if it can be purchased for $421.24.

Suppose that the machine that promises this income stream can actually be purchased for $432.94. Will it be a profitable investment? All we know so far is that if we buy the machine for $432.94 and if it produces net income of $100 per year for five years, we will have a 5 percent yield, or rate of return, on the funds invested in the machine over the five years. Is a prospective rate of return of 5 percent high enough to induce the business person to make the investment expenditure? It is here that the market rate of interest enters the picture. If the firm must borrow the funds to purchase the machine and the rate it must pay for such funds exceeds 5 percent, the interest rate then exceeds the expected rate of return, and the investment would be unprofitable. If, on the other hand, the market rate of interest is below 5 percent, the investment would be profitable. If the firm need not borrow but has on hand its own funds, the comparison is between the rate of interest it could earn by simply lending these funds in the market at interest and the rate of return it expects from an investment of these funds in the machine. If the rate of interest it can secure by lending these funds exceeds 5 percent, the firm will be wiser to lend than to invest in the machine. If the rate at which it can lend funds in the market is less than 5 percent, it will be wiser to invest in the machine than to lend the funds. As a general rule, *it pays to invest in a capital good if the rate of return expected from that capital good over its life exceeds the current market rate of interest.* This "rule," however, is subject to a number of qualifications, some of which we will consider later.

Investment Equations

All this may also be expressed in terms of the general equation for the present value of a future income stream. To develop this equation, we start with the ordinary equation for the growth of a sum at compound interest. Suppose that the market rate of interest is 5 percent. If one lends $95.24 at 5 percent for one year, he will receive $100 at the end of the year.

$$\$95.24 + (0.05 \times \$95.24) = \$95.24(1 + 0.05)$$
$$= \$100$$

In general, a lender will get back at the end of one year

third year $288.61. Withdrawing another $100 from investment, he invests $188.61 for the fourth year, on which he receives $7.54 in interest, making his investment at the end of the fourth year $196.15. Withdrawing another $100 from investment, he invests $96.15 for the fifth year, on which he receives $3.85 in interest, making his investment at the end of the fifth year $100, which amount he withdraws to close out the investment.

$$P_1 = P_0(1 + r)$$

where r is the rate of interest and P_0 is the amount lent. If he lends the amount P_1, which equals $P_0(1 + r)$, for the second year, he will get back at the end of the second year

$$P_2 = P_0(1 + r)(1 + r) = P_0(1 + r)^2$$

The rule is thus that a sum of P_0 lent at interest rate r for n years will grow at the end of n years to

$$P_n = P_0(1 + r)^n$$

We have seen that the process of discounting a future sum to the present is the reverse of the process of accumulation. What is the present value of \$100 receivable one year from today when the market rate of interest is 5 percent? We are given P_1 of \$100 and r of 0.05 and wish to find P_0. Since $P_1 = P_0(1 + r)$, it follows that

$$P_0 = \frac{P_1}{(1 + r)} = \frac{\$100}{(1 + 0.05)} = \$95.24$$

What is the present value of \$100 receivable two years from today? Since $P_2 = P_0(1 + r)^2$, we have

$$P_0 = \frac{P_2}{(1 + r)^2} = \frac{\$100}{(1 + 0.05)^2} = \$90.70$$

In general, we can find the present value of any series of future sums by discounting each portion of that series back to the present by the market rate of interest. Using V for present value and $R_1, R_2, \ldots, R_n$ for the series of sums, we have

$$\boxed{V = \frac{R_1}{(1 + r)} + \frac{R_2}{(1 + r)^2} + \cdots + \frac{R_n}{(1 + r)^n}}$$

In the case of the machine that is expected to produce a stream of income of \$100 at the end of each year for a period of five years, the present value of this income stream discounted at 5 percent is

$$V = \frac{\$100}{(1 + 0.05)} + \frac{\$100}{(1 + 0.05)^2} + \frac{\$100}{(1 + 0.05)^3} + \frac{\$100}{(1 + 0.05)^4} + \frac{\$100}{(1 + 0.05)^5}$$
$$= \$95.24 + \$90.70 + \$86.38 + \$82.27 + \$78.35$$
$$= \$432.94$$

These are the same results shown in the 5-percent column of Table 10-1.[5] If the actual market rate of interest were higher or lower than 5 percent, the present value of the income stream generated by the machine would be smaller or larger than \$432.94—for example, \$421.24 for a 6-percent market rate and \$445.18 for a 4-percent rate as shown in Table 10-1.

The Marginal Efficiency of Capital

To find the present value figure above, we started out with a given income stream and with the rate of interest that was assumed to be the actual market rate at the time. Or given $R_1, R_2, \ldots, R_n$, and r, we could find V. Suppose now that we are given the purchase price or cost of a capital good, designated as C, and the income stream to be produced by that good, designated as before. What we want to find in this case is the rate of interest that will make the present value of the income stream produced by the capital good just equal to the cost of the capital good. In terms of the following equation, we want to find the value of the rate i for given values of C and $R_1, R_2, \ldots, R_n$.

$$C = \frac{R_1}{(1 + i)} + \frac{R_2}{(1 + i)^2} + \cdots + \frac{R_n}{(1 + i)^n}$$

Note that the rate we find in this way is designated i to distinguish it from the market rate of interest r. The relationship between i and r will be examined below.

For a numerical illustration, assume a capital good whose cost is \$427.02 that is expected to produce an income stream of \$100 at the end of each year for a period of five years. Substituting these figures in the equation just above gives us

[5] The present value of a dollar discounted at various rates and for various time periods may be found from tables provided in financial handbooks and finance textbooks. See, for example, *Comprehensive Bond Values Table*, Financial Publishing Co., 1958 and H. Bierman, Jr. and S. Smidt, *The Capital Budgeting Decision*, 4th ed., Macmillan, 1975.

$$\$427.02 = \frac{\$100}{(1+i)} + \frac{\$100}{(1+i)^2} + \frac{\$100}{(1+i)^3} + \frac{\$100}{(1+i)^4} + \frac{\$100}{(1+i)^5}$$

Solving will give us 5.5 percent as the value for i.

We could substitute any other possible price for the capital good and solve for i on the basis of that price. For a given income stream, the higher the price of the good, the lower will be the value found for i; the lower the price of the good, the higher will be the value found for i. For example, in the case illustrated above, if the good cost \$485.35, i would be a mere 1 percent, but if it were "bargain priced" at \$299.00, we would find the value of i to be 20 percent.

In solving for the rate that will make the present value of the returns from a capital asset during its life just equal to its cost, we have found the rate that Keynes called the *marginal efficiency of capital*. This term has become part of the language of the theory of capital and investment and is one that we will use through the remainder of this book to indicate the rate of return expected from a capital asset.

In the manner described above, we may compute the marginal efficiency of capital, MEC or i, for any capital good once we are given its cost and the stream of income expected from it. By comparing the MEC with the current market rate of interest, r, we can tell at once whether the contemplated investment promises to be profitable or unprofitable. Thus, in the illustration above, given that the capital good in question costs \$427.02, we found its MEC to be 5.5 percent. If the rate of interest were 6 percent, the investment promises to be unprofitable, but if the rate of interest were 5 percent, it would be considered profitable. Furthermore, simply by finding the difference between the MEC and r, we have the net rate of return expected on the capital asset after allowance for all costs, including the interest cost on the funds tied up in the capital good over its life and the depreciation cost of the asset itself. Thus, with the MEC at 5.5 percent and r at 5 percent, purchase of the capital good promises a net return of 0.5 percent over and above all costs; with r at 6 percent, a net return of -0.5 percent.

Because the MEC and r are both percentages, they are sometimes confused or, even worse, identified as the same thing. It is essential to see that the two percentages are distinctly different, and that the business person's estimate of the MEC for any capital good depends in no way on r. Once the MEC has been estimated, the profitability of the capital good in question can be gauged only by comparing its MEC with r, but this is a step altogether distinct from that of estimating the MEC itself. The level of r determines whether the good will be purchased, once its MEC is given, but r does not in any way determine the MEC of that good.

We may now note how all three elements—the income flow expected from a capital good, the supply price of that good, and the market rate of interest—fit together. An improvement in the business outlook that causes the business person to revise upward his estimate of the expected income flow from a capital good will, given an unchanged price for the good, raise the MEC of that good. Alternatively, if there is no change in the income flow expected from the good, a drop in its price will raise its MEC. A fall in r will not affect the MEC of that good, but if $r >$ MEC before the fall and if MEC $> r$ after the fall, the purchase of the capital good, which previously appeared unprofitable, will now appear profitable. A downward revision in the expected income flow from the good, a rise in its price, or a rise in the market rate of interest will all work in the opposite direction—the direction of decreasing the expected profitability of the good or even turning it into expected loss.

 market rate - r

STOCK OF CAPITAL AND THE RATE OF INVESTMENT

At any point in time, a firm is confronted with a long list of possible investment projects. Apart from projects that arise as a result of changing technology, there are projects such as expansion of its existing factory building, construction of a new and larger building, purchase of more of the same equipment to expand production of existing lines, purchase of different equipment to produce a new line of goods, and purchase of trucks to handle its own deliveries. If the firm already has or can borrow the necessary funds, the investment expenditures for every possible project will or will not be made depending upon the MEC of each and the then current level of *r*. After estimating the MEC of each of the diverse investment projects it might undertake, the firm could prepare a schedule like Figure 10-2 to show the possibilities.

Figure 10-2 assumes that at a selected point in time the firm's stock of capital net of depreciation is valued at $500,000. If the market rate of interest is 6 percent, the first two projects in the schedule promise to be profitable. These projects require net investment expenditures of $125,000, and they will increase the firm's capital stock to $625,000. If the market rate of interest were 3 percent, the third and fourth projects would also be profitable. Net investment expenditures of another $75,000 would raise the firm's stock of capital to $700,000. Once the stock of capital has been expanded to the level at which the MEC of the last project has dropped to the level of *r*, the firm has that stock of capital which is for it the profit-maximizing or desired stock. No further change is to be expected, barring the appearance of new projects with MEC above *r* or barring a fall in *r* itself. The firm's net investment in each succeeding period will be zero; its gross investment will be whatever is required for replacement purposes only.

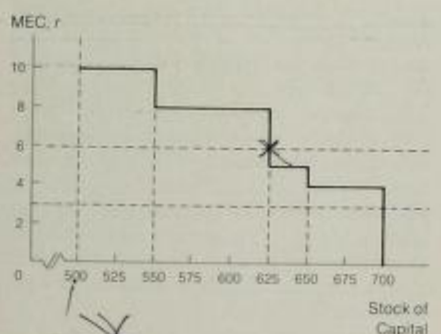


FIGURE 10-2
Schedule of the marginal efficiency of capital for a firm

What is true of one firm is true of others—management of every profit-conscious firm is continuously on the outlook for investment opportunities that promise to improve the appearance of the profit and loss statement over the years. If each firm in the economy were to prepare a schedule of the kind shown in Figure 10-2 and if all these schedules were added together, we would have an aggregate schedule of the MEC such as that shown in Part A of Figure 10-3.[6] The "lumpiness" of the invest-

[6]In a strict sense, the aggregate schedule cannot be derived in so simple a manner. A drop in the rate of interest from 6 to 4.5 percent in Figure 10-2 may call forth more than $25,000 of net investment by the firm in question. Since the same drop in the interest rate will trigger investment by other firms, the resultant increase in aggregate spending may so improve the outlook of this firm as to raise the MEC of its other prospective projects above what they were before. In general, when we allow for this interdependence, the effect is to make the individual firm's schedule more elastic than otherwise. The determination of the aggregate schedule, however, becomes a problem in general equilibrium analysis, which allows for the interdependence between each firm's schedule and those of all other firms.

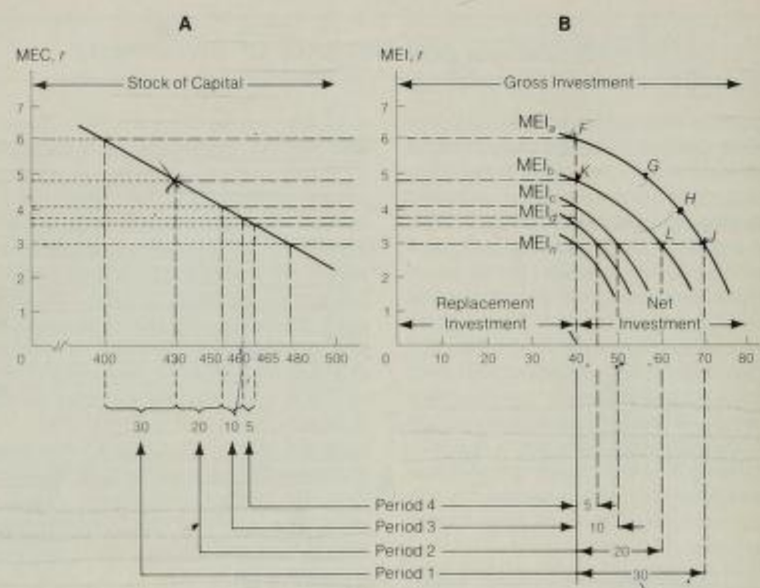


FIGURE 10-3
The process of capital accumulation in response to a fall in the rate of interest

ment projects available to the individual firm that produces the stair-like curve of Figure 10-2 evens out in aggregation to produce the smooth curve of Part A of Figure 10-3.

At any point in time, all firms combined have some existing aggregate stock of capital, say $400 billion. At the same point in time, there is some particular market rate of interest, say 6 percent (here assuming that a single rate prevails for all borrowers and for loans of all maturities and for all purposes). With the aggregate MEC schedule as given in Part A of Figure 10-3 and with *r* at 6 percent, any expansion in the aggregate stock of capital beyond the existing $400 billion level promises a net return of less than 6 percent, or a return below the market rate of interest. Under these conditions, business as a whole may be said to have that stock of capital which maximizes profits; there is no reason to expand or contract the capital stock, and so there is no net investment or disinvestment. For some individual firms there may be net investment, but for others there will be an equal amount of disinvestment, making net investment zero for all firms combined. Under these conditions, there can be no net investment for all firms combined until there is either a drop in the rate of interest that produces a movement down the MEC schedule to a higher stock of capital, or a shift to the right of the MEC schedule that will, with an unchanged rate of interest, make the stock of capital consistent with profit maximization larger than it was. Let us examine in turn the effect of a change in the rate of interest and a shift in the MEC schedule.

A Fall in the Rate of Interest

In our example, we have assumed that with a capital stock of $400 billion, MEC is 6 percent. If r is also 6 percent, this would be the profit-maximizing stock of capital, and net investment spending would be zero. Assume now a drop in r from 6 to 3 percent.[7] At this lower rate, the profit-maximizing stock of capital as given by the MEC schedule of Figure 10-3 is $480 billion. This means that $80 billion of net investment expenditures is needed to raise the capital stock to the higher, desired level. How much will all firms combined actually invest per time period? And how long will it take these firms to build up the capital stock from the $400 billion to the $480 billion level?

In considering the individual firm, we were able to bypass this kind of question. When a firm decides to increase its stock of capital, it can conceivably accomplish this virtually overnight. If the capital goods required are goods held in inventory by firms in the capital goods industries, the firm can move up to the higher capital stock desired as quickly as these goods can be delivered. If new construction is required, the firm can move up to the higher capital stock as quickly as this construction can be completed. Furthermore, it is unlikely that its investment expenditures would exert more than a negligible upward pressure on the supply prices of the capital goods in question.

For firms as a group, we cannot, however, bypass this kind of question. To expand the aggregate stock of capital from $400 to $480 billion must take time. Although some firms may increase their stock of capital immediately by simply buying capital goods from other firms, there will be no increase in the stock of capital goods held by all firms combined unless more capital goods are produced, and the rate at which these goods can be produced is limited by the existing productive capacity of the capital goods industries.[8] Furthermore, as all firms combined step up their investment expenditures, the capital goods industries, in expanding output to meet this higher demand, will at some point experience rising marginal costs that may lead to higher prices for capital goods. These rising prices will in turn slow the rate of investment spending and thus lengthen the time period required to effect any given increase in the aggregate stock of capital.

Let us assume that the capacity output of the capital goods industries is $80 billion per time period, valuing capital goods at the prices in effect when their output is just the amount needed for replacement requirements.[9] With an existing capital stock of $400 billion, replacement requirements are $40 billion per time period, if we assume that 10 percent of these goods wear out in each time period. This means that the capital goods industries, operating at capacity output, can supply a net addition to the capital stock of $40 billion per time period. In other words, it will take two time periods *at capacity output* to raise the capital stock from $400 to $480 billion. Will the rate of net investment spending be sufficient to push the capital goods industries to capacity output? To find out what net investment expenditures will be per time period for all firms combined, turn to Figure 10-3.[10]

[7] This drastic change is unrealistic, but it is chosen because it will show clearly in chart form the results that follow a fall in the interest rate.

[8] Of course, the productive capacity of the capital goods industries may grow over time as these industries expand through net investment, but, for simplicity, we here assume no such change, or constant productive capacity in the capital goods industries.

[9] Since the MEC schedule will shift with every change in the supply price of capital goods, to produce the stable MEC schedule necessary to the following analysis requires that we assume that the MEC schedule is based on the fixed supply price for capital goods that is in effect when output of capital goods is, say, the amount necessary for replacement purposes only. This amount can reasonably be assumed to be stable in the short run.

[10] This figure is adapted from G. Ackley, *Macroeconomic Theory*, Macmillan, 1961. See pp. 481–85. For alternative treatments, see A. P. Lerner, "On Some Recent Developments in Capital Theory," in *American Economic Review, Proceedings*, May 1965, pp. 284–95, *The Economics of Control*, Macmillan, 1944, pp. 330–38, and R. L. Crouch, *Macroeconomics*, Harcourt Brace Jovanovich, 1972, pp. 66–86.

Part B of Figure 10-3 introduces the schedule of the *marginal efficiency of investment*, or MEI schedule. This schedule indicates the rate of investment spending per time period at each possible market rate of interst. Separate MEI schedules are shown for a number of different levels of the stock of capital. With the MEC schedule as given in Part A, with r at 6 percent, and with the capital stock at \$400 billion, the actual capital stock is the profit-maximizing stock, and net investment is zero. Therefore, we know that a schedule relating investment expenditures to the rate of interest must show zero net investment at an interest rate of 6 percent, which the MEI schedule labeled MEI_a does show (at point F). If, as we have assumed, r falls to 3 percent, net investment expenditures will appear. In the first time period after the drop in r, curve MEI_a tells us that net investment will be \$30 billion (or gross investment will be \$70 billion), even though the capacity of the capital goods industries permits \$40 billion of net investment (or \$80 billion of gross investment) per time period. Investment spending is checked at a rate short of the capacity output of the capital goods industries by the rise in prices of these goods, which occurs as their rate of output is expanded in response to investment demand. *It is specifically the upward-sloping supply curve of capital goods that produces the downward-sloping MEI curve.* Furthermore, the more sharply the supply curve of capital goods slopes upward, the more sharply the MEI curve slopes downward.[11]

We can see more clearly why net investment in the first period following the fall in r will be exactly \$30 billion by starting at point F on the MEI_a schedule and following this schedule down to point G. At G the MEI is 5 percent because the rise in net investment spending from a rate of zero to a rate of about \$15 billion has so raised the prices of capital goods as to reduce to 5 percent the rate of return on investment in these goods. This, however, is still an MEI above r, which is 3 percent, so a higher rate of net investment spending is warranted. At point H the MEI has fallen to 4 percent because the still higher rate of net investment spending, now about \$25 billion, has pushed the prices of capital goods to the still higher level at which the rate of return on investment in them is reduced to 4 percent. The MEI is still above r, so a still higher rate of net investment spending is warranted. At point J, the MEI is 3 percent, again for the reason that the still higher rate of net investment spending, now \$30 billion, has pushed the prices of capital goods to the still higher level at which the rate of return on investment in them is reduced to 3 percent. It may now be seen that to push the rate of net investment spending beyond \$30 billion would reduce the MEI still further, or below 3 percent and so below r. Accordingly, net investment will be at the \$30 billion rate, no higher and no lower, in the first time period following the drop in r.

From the beginning to the end of this first time period, with net investment of \$30 billion during the period, the capital stock will have risen to \$430 billion from its beginning level of \$400 billion. As shown in Part A, this increase of \$30 billion in the capital stock reduces the MEC to about 5 percent (actually 4.87) from its previous level of 6. Since r is still 3 percent, a further growth in the capital stock is called for. What will be the amount by which it grows in the second period, or what will be the rate of net investment spending in the second period? This depends on the MEI schedule. With MEC now at 5 percent, we know that the new MEI schedule, relating investment expenditures to the rate of interest, must show zero net investment at an interest rate of 5 percent (point K). Thus, MEI_b, the new schedule, lies below MEI_a because of the increase in the stock of capital in the first time period; the whole MEI

[11]Suppose the supply curve were perfectly elastic up to the capacity output. Then the MEI curve, MEI_a, instead of sloping downward, would be perfectly horizontal, or perfectly elastic, up to gross investment of \$80 billion. The same would be true for the other MEI curves that are explained below. An MEI curve perfectly elastic over an unlimited range will be encountered in the following chapter's discussion of the simple acceleration principle, which assumes an MEI curve of that kind.

schedule must fall to a lower level with each movement to a lower point on the given MEC schedule. MEI_b slopes downward for the same reason that MEI_a sloped downward—the rising supply price of capital goods as the rate of output of these goods is expanded in response to investment spending.

The rate of net investment spending in the second period is determined in the same way as the rate in the first period was—it will be that rate which reduces the MEI to equality with r. Schedule MEI_b shows that net investment for the second period will be \$20 billion (point L). Note that this is below the \$30 billion found in the first time period. Given that the prices of capital goods rise with their rate of output, as soon as net investment reaches the rate of \$20 billion per time period, the rise in prices of capital goods reduces the MEI from 5 to 3 percent, or to equality with r. In the first period, only when net investment reached the \$30 billion rate was the rise in prices of capital goods sufficient to produce the greater drop in MEI from 6 to 3 percent, or to equality with r. In other words, the greater spread between MEI and r at the beginning of the first period permitted the higher rate of investment spending in that period.

Net investment spending of \$20 billion during the second period raises the stock of capital to \$450 billion from its level of \$430 at the beginning of the second period. The increase in the stock of capital reduces the MEC further to about 4 percent (actually 4.12), which produces the new, lower MEI schedule MEI_c. The rate of net investment spending in the third period as given by this schedule is \$10 billion. This again raises the stock of capital, now from \$450 to \$460 billion. This in turn reduces the MEC, now below 4 percent, and creates the new, still lower MEI schedule MEI_d. The rate of net investment spending in the fourth period is then \$5 billion. With no shift in the MEC schedule and with no further fall in r, net investment spending, lower in each succeeding period, eventually raises the stock of capital to \$480 billion, at which level the MEC equals r. The actual stock of capital is now the profit-maximizing stock of capital for the interest rate of 3 percent. With the capital stock at \$480 billion, the relevant MEI schedule is MEI_n, which shows net investment to be zero and gross investment equal to replacement investment per time period. We have reached a new equilibrium, which will be upset only by a shift in the MEC schedule or by a change in the market rate of interest.

To summarize: Given an unshifting MEC schedule, the appearance of a positive rate of net investment is seen to depend in the first instance on the size of the stock of capital and the rate of interest. Given the stock of capital, we know MEC; and given r we know whether $MEC > r$, $MEC < r$, or $MEC = r$. A prerequisite to net investment is that $MEC > r$. If we have this situation, an increase in capital stock is called for, and this increase can come about only through net investment spending. The rate of net investment spending per time period depends on how steep the downward slope of the MEI schedule is (or more correctly its elasticity), and this in turn depends on how steep the upward slope (or the elasticity) of the supply curve of capital goods is. If the supply curve slopes sharply upward, the rate of investment spending will fall sharply downward with respect to the rate of interest. In any event, the capital stock will grow to the new profit-maximizing level, but its rate of growth will be slower the steeper the MEI schedule.

A Shift in the MEC Schedule[12]

Just as a drop in the rate of interest with no shift in the MEC schedule will raise the profit-maximizing stock of capital from its previous

[12] Although the intent here is simply to distinguish between a change in the rate of interest and a shift in the MEC schedule, it should be noted that these two changes are not independent of each other. An increase in investment spending resulting from a fall in the rate of interest raises the level of income, which in turn (via the acceleration principle to be examined in Chapter 11) tends to cause a shift to the right in the MEC schedule. A movement along the schedule may, in other words, set into motion forces that cause a shift in the schedule.

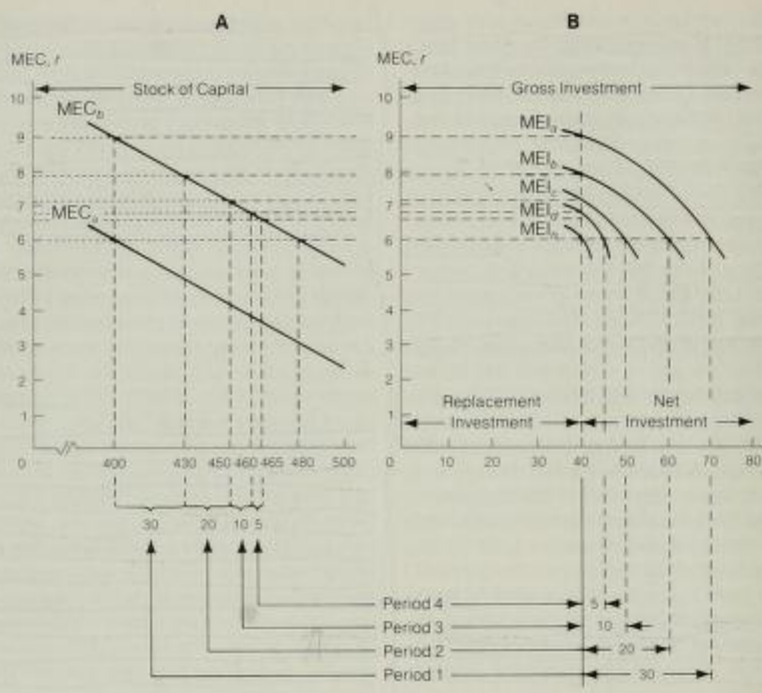


FIGURE 10-4

The process of capital accumulation in response to a shift in the schedule of the marginal efficiency of capital

level, so an upward shift in the MEC schedule with no change in the rate of interest will have the same result. Schedule MEC_a of Figure 10-4 is identical with the MEC schedule shown in Figure 10-3. In both cases, assuming that *r* is 6 percent and the actual capital stock is \$400 billion, the actual stock is the profit-maximizing stock, and net investment spending is zero. If the MEC schedule now shifts upward from MEC_a to MEC_b, the profit-maximizing stock of capital becomes \$480 billion. Net investment now appears, because, with the capital stock at \$400 billion, the MEC has risen from 6 to 9 percent. In other words, due to a rise in the expected income flow from capital goods, which is assumed here to be the cause of the upward shift in the MEC schedule, the rate of return expected from the first increment to the stock of capital over its existing level is 9 percent. Since *r* is still 6 percent, the MEC now exceeds *r* by 3 percent. Note that this is the same relationship between MEC and *r* as was given in Figure 10-3, the only difference being that here the 3 percent spread between MEC and *r* is brought about by a rise in MEC from 6 to 9 percent, whereas in Figure 10-3 the same spread was brought about by a drop in *r* from 6 to 3 percent.

As in the previous illustration, net investment expenditures of $80 billion are required to raise the capital stock to its new profit-maximizing level. The rate of net investment in each time period is the same in both illustrations, as is the number of time periods required to increase the capital stock by $80 billion. Part B of Figure 10-4 differs from Part B of Figure 10-3 only in being higher above the horizontal axis. In Figure 10-3, investment in the first period after the fall in r was at the rate that reduced the MEI from 6 to 3 percent; in Figure 10-4, investment in the first period after the shift upward in the MEC schedule is at the rate that reduces the MEI from 9 to 6 percent. *The significant factor is the spread between MEI and r rather than the absolute level of either.* Since this spread is the same in both of our illustrations, the rate of net investment in each period will be the same.

It is, of course, quite possible that shifts in the MEC schedule could occur at the same time as changes in the market rate of interest. For example, the upward shift in the MEC schedule of Figure 10-4 could be accompanied by a fall in the interest rate like that shown in Figure 10-3. If this happened, the new profit-maximizing capital stock would be greater than $480 billion. If there were a rise in the rate of interest, it might be sufficient to offset the rise in the profit-maximizing capital stock that otherwise would result from an upward shift in the MEC schedule. Conversely, a fall in the rate of interest could be such as to offset the decrease in the profit-maximizing capital stock that otherwise would result from a downward shift in the MEC schedule.

Unless these combinations of changes are exactly offsetting, however, they will change the profit-maximizing capital stock. Given such a change, the net investment process as described is that by which the economy moves to the higher profit-maximizing capital stock, whatever it may be and whatever may have caused it to change.[13]

A Summary Formulation

The portion of aggregate spending that is accounted for by the business sector of the economy is measured by the investment expenditures of business for newly produced capital goods. To explain aggregate spending for any time period, we must concentrate on business investment expenditures for that time period rather than on the actual stock of capital held by business at any point within that time period. Net investment expenditures for any time period are the means by which a change in the capital stock is effected. In other words, net investment, which is zero when the actual capital stock equals the profit-maximizing capital stock, becomes positive when the profit-maximizing stock exceeds the actual stock. A prerequisite to the appearance of net investment expenditures is a rise in the profit-maximizing capital stock.

What produces such a change in the profit-maximizing capital stock? Once such a change occurs, the rate of net investment expenditures determines the time necessary to raise the capital stock to its profit-maximizing level. What determines this rate? In the preceding pages we have attempted to lay the conceptual groundwork for answering these two questions. Figure 10-5 presents in schematic form the various factors we have introduced so far. If the difference between the actual and profit-maximizing capital stock is zero, the rate of net investment

[13]Since, apart from several years during the Great Depression, the profit-maximizing capital stock has grown uninterruptedly over the years, we have not here specifically entered into the process by which the capital stock would be reduced over time below the actual capital stock at any point in time. In brief, if the movement were, say, from an actual capital stock of $480 billion to a desired capital stock of $400 billion, gross investment would drop all the way to zero, and net investment would accordingly become negative. The maximum possible rate of negative net investment, or of disinvestment, is set by the rate at which the capital stock is used up. If, as earlier assumed, this were simply 10 percent per time period, it would then take a little less than two time periods to reduce the actual stock from $480 billion to the desired stock of $400 billion. Since no capital goods are being purchased, the shape of the supply curve of capital goods and so the shape of the MEI schedule have nothing to do with the maximum disinvestment rate.

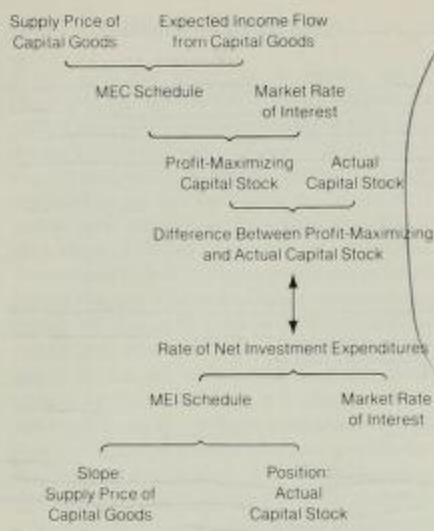


FIGURE 10-5
From the stock of capital to the rate of investment

expenditures will be zero. In this case, the portion of the chart lying below the vertical arrow is not really relevant; it becomes relevant only when there is a change somewhere above the arrow that produces a difference between the actual and the profit-maximizing capital stock. Whether such a difference will appear depends on the relationship between the MEC schedule and the market rate of interest. The MEC schedule in turn depends on the supply price of capital goods and the expected income flow from such goods.[14] If there is a change in any of these factors that is sufficient to produce a difference between the actual and the profit-maximizing capital stock, then the lower portion of the chart gives us the factors that determine the rate of net investment and thus the time required to raise the actual capital stock to the profit-maximizing level. Here we recall that the rate of net investment expenditures depends on the relationship between the rate of interest and the MEI schedule, that the overall slope of the MEI schedule depends in turn on the supply price of capital goods, and that its position vis-à-vis the axes depends on the actual stock of capital.

[14] As was noted earlier, in order to derive an MEC schedule, a particular price level of capital goods must be assumed. This means assuming the price level to be given at some point along the supply price schedule. For each different price level selected, there will be a different MEC schedule; but for any one price level, there will be a unique MEC schedule. To derive the MEI schedule, on the other hand, the complete supply price schedule is employed; it determines the overall slope of the MEI schedule.

The MEI Schedule: Long Run and Short Run

So far we have discussed one factor that influences the rate of net investment spending—the growth in the stock of capital, represented graphically by a movement down the MEC schedule. This factor becomes important only in the long run; then the growth in the stock of capital is large enough, relative to the preexisting stock of capital, to cause an appreciable movement down the MEC schedule. For this reason, it is necessary that we now distinguish between the MEI schedule in the long run and the MEI schedule in the short run.

The process illustrated by Figures 10-3 and 10-4 takes place over the long run. Given a one-time drop in r and no shift in the MEC schedule, Figure 10-3 showed how, over the course of the first few of an endless series of time periods, the stock of capital would grow toward the new profit-maximizing stock as a result of each period's net investment expenditures. Figure 10-4 showed the same process for a one-time shift in the MEC schedule with no change in r.

In both cases we deliberately exaggerated the changes involved to clarify the process. The net investment expenditures of each period were sizable relative to the capital stock at the beginning of each period; this in turn meant

that the resulting change in the stock of capital in each period would push down the MEI schedule perceptibly in the next period until that schedule eventually reached the level at which net investment expenditures became zero. This explanation must now be qualified to describe the MEI schedule in the short run. In the short run, the addition to the stock of capital resulting from net investment expenditures will be insignificant relative to the large existing capital stock. Therefore, the downward shift in the MEI schedule shown in Parts B of Figures 10-3 and 10-4 will be correspondingly insignificant. As a matter of fact, a short-run period in the present context is usually defined as a time interval of such length that the changes in capital stock, relative to the size of the capital stock before the changes, are too small to influence the level of net investment expenditures.

For short-run analysis, it is then possible, without appreciable error, to abstract from the effect of changes in the capital stock on the position of the MEI schedule. In terms of Figure 10-3, when *r* falls from 6 to 3 percent, the result is net investment of $30 billion in the first period following the fall in *r*. However, instead of the long-run results described by Figure 10-3, we may assume that the preexisting stock of capital is so large that the addition of $30 billion to that stock does not cause a perceptible movement down the MEC schedule. With no appreciable movement along the MEC schedule, the MEI schedule in the second period is virtually in the same position as in the first period. This means that the rate of investment expenditures in the second period and in subsequent periods that make up the short run may remain virtually the same as that of the first period. This, it may be noted, is precisely what Keynes assumed in his *General Theory*.

In Chapter 21, on economic growth, we will be concerned with long-run analysis; we will have to look both ways in the manner originally described in connection with Figures 10-3 and 10-4. In the long run, net investment expenditures do move the economy along the MEC schedule. With the MEC schedule sloping downward, the result must be a downward shift of the MEI schedule and, assuming an unchanging supply curve for capital goods, a decrease in the rate of net investment expenditures. However, in this part of the book, we need look only one way—a difference between the actual and the profit-maximizing capital stock affects net investment expenditures, but these expenditures *in the short run* do not appreciably affect the economy's position along the MEC schedule.

A CONCLUDING NOTE

Our concern in this chapter has been essentially with the meaning of capital and investment and with the mechanics of the relationships between the stock of capital and the rate of investment. If at any point in time the actual stock equals the profit-maximizing stock, an excess of the latter over the former will appear subsequently only if the market rate of interest falls or if the MEC schedule shifts upward. Either change will give rise to a positive rate of net investment spending. For short-run analysis, we may assume that the indicated rate of net investment spending may continue unchanged. However, in the long run the rise in the stock of capital resulting from net investment spending will depress the MEI schedule and with it the rate of net investment spending. This long-run result assumes that the growth in the stock of capital produces a movement along an unshifting MEC schedule, a schedule that necessarily slopes downward. In reality, however, it is possible that, long before net investment

spending raises the actual stock of capital to the profit-maximizing level indicated by a given market rate of interest and a given MEC schedule, this schedule will have shifted and from this shift will follow a shift in the MEI schedule and from it a change in the rate of net investment spending.

Now that we understand the mechanics by which changes in the profit-maximizing capital stock are translated into changes in the rate of investment, we are free to concentrate on the factors whose changes produce shifts in or movements along the MEC schedule and thereby produce changes in the profit-maximizing capital stock. We turn to this in the following two chapters.

chapter

11 22

Investment Spending: The Profits and Accelerator Theories

What was said over a decade ago by two prominent students of investment spending is probably equally true today: "The theory and measurement of investment behavior is one of the most controversial areas of professional economic study. . . . the subject is inherently difficult and complex."[1] In view of these characteristics, it is not surprising that there is a plethora of theories to explain investment demand. We will not attempt to present here all the major theories, but will limit our attention to the two that were noted at the beginning of the preceding chapter: the profits theory and the accelerator theory.[2]

In the simplified version that we will develop, it may even be incorrect to classify the profits theory as one of the major theories. However, a simple profits theory is the one that most laymen think of first (and often the only one they think of) and is worth consideration as long as the appropriate criticism is included. As for the accelerator theory, practically all economists assign a place of major importance to it, at least in one of its more refined versions, although many disagree with those economists who hold that this theory by itself provides an adequate explanation of investment demand.

In terms of the apparatus developed in the preceding chapter, we can classify all theories of investment or all of the individual factors influencing investment into those that exert their influence by producing a movement along an existing MEC schedule and those that exert their influence by producing a shift in the schedule itself. The profits theory and the accelerator theory are both of the latter type. In the profits theory the desired capital stock is a function of profits. Accordingly, a rise in profits causes a rightward shift in the MEC schedule, indicating a larger desired capital stock at each rate of interest. In the accelerator theory the desired capital stock is a function of the economy's output. Accordingly, a rise in output

[1] J. R. Meyer and R. R. Glauber, *Investment Decisions, Economic Forecasting, and Public Policy*, Harvard Business School, 1964, p. 1.

[2] The number of theories obviously goes up from here. A five-way classification of theories has been employed in one empirical study: (1) neoclassical I, (2) neoclassical II (the difference being in the treatment of capital gains and losses on capital goods), (3) accelerator, (4) expected profits, and (5) liquidity. See D. W. Jorgenson and C. D. Siebert, "A Comparison of Alternative Theories of Corporate Investment Behavior," *American Economic Review*, September 1968, pp. 681–712. Some authors have combined in a single model the basic idea of more than one of the five theories noted. See, for example, R. W. Resek "Investment by Manufacturing Firms: A Quarterly Time Series Analysis of Industry Data," *Review of Economics and Statistics*, August 1966, pp. 322–33, and M. K. Evans, "A Study of Industry Investment Decisions," *Review of Economics and Statistics*, May 1967, pp. 151–64.

causes a rightward shift in the MEC schedule, indicating a larger desired capital stock at each rate of interest. According to these two theories, a decrease in profits or in the level of output will, of course, produce leftward shifts in the MEC curve.

As we saw in the preceding chapter, a rise in the desired capital stock and thereby a rise in investment expenditures may also occur without a shift in the MEC curve. If there is a reduction in the market rate of interest, there is then a movement down along the existing MEC curve to a larger desired capital stock. All the factors that affect the market interest rate thereby become factors affecting the desired capital stock and through it investment expenditures. How important these may be depends in large part on the elasticity of the MEC schedule. The more elastic it is, the more important are changes in the interest rate as an influence on the rate of investment expenditures. While there is some difference of opinion among economists as to the elasticity of the schedule, there is virtual unanimity of opinion that shifts in the schedule are much more important in explaining the observed changes in investment expenditures than the movements along a given schedule that result from changes in the interest rate. In the following chapter we will turn to the questions of the elasticity of the MEC schedule and the importance or unimportance of the rate of interest and the role of finance as influences on investment spending. Meanwhile, in the remainder of this chapter we will, as noted previously, consider two of the various theories that have been advanced to explain why the MEC schedule shifts as it does—the profits theory and the accelerator theory.

THE PROFITS THEORY[3]

At first glance, it may seem that the profits theory is the only possible theory of investment because, in an economy in which profit maximization is the primary goal of business, firms will rarely engage in investment expenditures for capital assets that are not expected to add to their profits (or reduce their losses). Additional units of plant and equipment will be acquired only if each addition can be expected to bring with it an increment to total profits. In terms of the framework of the preceding chapter, this will be the case as long as the MEC exceeds the rate of interest.

This basically incontrovertible proposition is not, however, what is meant by the profits theory of investment. The profits theory holds that the amount of investment spending depends on the amount of profits that firms are making. This is distinctly different from the statement of the preceding paragraph. We may all agree that investment spending will be undertaken only in the expectation that it will be profitable, but we cannot find in this any explanation of the actual volume of investment spending that will take place in any time period. We are not likely to agree so unanimously that a higher dollar amount of profits being earned will give rise to a higher dollar amount of investment, but we will unanimously agree that this is a hypothesis deserving of investigation as a possible explanation of what determines investment spending.

The crucial question may now be posed: What is the basis for the theory that the level of

[3]Among the major studies that deal with the effect of profits on investment are the following: J. Tinbergen, *A Method and Its Application to Investment Activity*, League of Nations, 1939; C. F. Roos, "The Demand for Investment Goods," in *American Economic Review, Papers and Proceedings*, May 1948, pp. 311–20; L. R. Klein, "Studies in Investment Behavior," in *Conference on Business Cycles*, Universities–National Bureau Committee for Economic Research, 1951; J. R. Meyer and E. Kuh, *The Investment Decision*, Harvard Univ. Press, 1957; and Yehuda Grunfeld, "The Determinants of Corporate Investment," in A. C. Harberger, ed., *The Demand for Durable Goods*, Univ. of Chicago Press, 1960.

investment expenditures depends on the level of realized profits? A key element in the answer is found in the fact that one of each firm's guides to the probable future level of its profits is simply the level of profits of the current period and of the recent past. An increase in the level of profits actually experienced over a period of time may well give rise to the expectation of a continuation of this movement in the future. In estimating the stream of net income, $R_1 + R_2 + R_3 + \cdots + R_n$, that any prospective investment project may generate over its life, business persons are likely to come up with higher dollar figures if these estimates are being made at a time when the absolute levels of their company profits and economy-wide profits are rising than they would under the opposite set of conditions. A high level of profits in this way tends to make for higher estimates of the MEC of various investment projects than would otherwise be the case. In terms of the graphic apparatus developed in the preceding chapter, this means an upward or rightward shift in the MEC schedule, which amounts to an increase in the profit-maximizing or desired capital stock. This in turn translates into a rise in the level of investment spending.

This, at least, is what lies at the heart of the profits theory. Plausible as it sounds, it is still not difficult to find fault with it. For one thing, it is not obvious that the firm's realized profits of this year or of the last few years can provide a measure of the profits of next year or of the next few years. A rise in profits currently realized may be the result of unexpected changes that are of a transitory nature. If there is no reason to expect this situation to continue, the mere appearance of such profits does not provide an incentive to invest. Or suppose that a rise in profits is not the result of an unexpected change but is the normal payoff on past investment spending. Although this does say that the past investment spending turned out satisfactorily, it does not necessarily follow that it gives the firm an incentive to invest further.

From this point of view, the profits theory of investment is as good an explanation of investment spending as realized profits are an indicator of all those future conditions that determine whether or not an investment will be profitable. The firm's current profits reflect the current conditions of demand for its products and the current conditions of supply for the various inputs it uses. However, they do not adequately reflect the many changes that may occur in supply and demand conditions in the future. Conditions will, of course, vary drastically from industry to industry and from firm to firm, but even firms operating in oligopolistic or near-monopolistic markets may not be able to expect that the demand that made possible the realized profits of the present and near past will do the same in the future. A satisfactory level of profits today reflects satisfactory demand conditions today, but management can hardly conclude from this that the increase in demand will be there tomorrow to provide the increase in profits necessary to justify investment by the firm today. The same is true for supply conditions. For example, future technological developments may substantially affect methods of production and business profits, but these changes in tomorrow's supply conditions are not reflected in today's profit figures.

These considerations are sufficient to convince us that firms must assuredly look beyond their realized profits in deciding on changes in their capital stock. On the other hand, another consideration leads us to continue to assign considerable importance to current profits. Higher current profits make possible the internal financing of a larger volume of investment, and the total amount of investment actually undertaken may depend in part on how much can be internally financed. Without the internal funds generated by profits, certain investment projects that are otherwise attractive will not be undertaken; business may not choose to go ahead if it can do so only by raising funds externally.[4] Beyond this, business may even go

[4] The broader question of the cost and supply of investment funds in general will be touched on in the following chapter.

ahead with investment projects that are distinctly marginal as long as internal funds are available for their financing. The alternative is the commitment of all such funds to extra dividends or additions to cash, bank account, or short-term security holdings, none of which may appear as attractive to management as the available investment projects.

To the extent, then, that some investment projects that would not otherwise be undertaken are undertaken because current profits provide the means of their financing, we have another reason why the level of profits, or more exactly the level of profits remaining after taxes and dividends, may be important in determining the level of investment spending.

Insofar as the level of investment spending is responsive to the level of profits, it may also be said to be responsive to the level of income. This simply recognizes that the aggregate profits earned by business vary directly with the level of the economy's income. For a simple illustration, suppose that the change in profits was always equal to one-fifth of each change in income and that the change in investment was always equal to three-fourths of each change in profits. There would then be a rigid link between a change in income and a change in investment. A \$10 billion rise in income would mean a \$1.5 billion rise in investment [\$1.5 billion = (1/5) × (3/4) × \$10 billion].

No such rigid link, of course, exists. The profits theory is based on a link between investment and profits, but it does not require a rigid link in order to be a worthwhile theory. The historical record also shows no rigid relationship between aggregate profits and the level of income. The two vary in the same direction but in nothing like a proportionate manner. The total amount of profits is obviously much lower when the economy is operating at the relatively low level of income that is found during recession than it is when the economy is enjoying the relatively high level of income that characterizes prosperous periods. However, as the income level rises and falls over the course of business fluctuations, profits typically vary more than proportionally with the change in income. For example, from 1975 I to 1977 I, the first two years of recovery from the 1973–75 recession, national income increased 24.5 percent while corporate profits shot up by 68.4 percent; during the recession itself, the changes were correspondingly different in the opposite direction.

What we then have are relationships in which the level of aggregate profits varies with the level of the economy's income and, according to the profits theory of investment, the desired capital stock varies with the level of aggregate profits. In the event that the level of profits indicates a desired capital stock that exceeds the actual capital stock, net investment occurs to correct the shortage of capital, a process examined in detail in the preceding chapter.

The Level of Profits, the Desired Capital Stock, and the Investment Function

We can tie together the analysis from the preceding chapter and the profits theory of investment introduced here with the help of Figure 11-1. We begin with the curve in Part A, which shows that total profits, designated by Z, vary directly with the income level, Y. In accordance with the profits theory of investment, for each level of profits shown in Part A there is a particular capital stock desired by business. However, we cannot identify that particular capital stock in Part B without also knowing the market rate of interest, because the model now before us includes the rate of interest as well as the level of profits as a determinant of the desired capital stock. We meet this problem in Part B by constructing a family of curves, each corresponding to a different interest rate. The upward slope of each of these curves shows that the desired capital stock measured on the horizontal axis varies directly with the level of profits measured on the vertical axis. However, for any

particular level of profits, the desired capital stock will be smaller the higher the rate of interest; therefore, the higher the rate of interest, the further to the left lies the curve relating desired K to realized Z. Curves for 8, 6, and 4 percent interest rates are shown in Part B.

We are now able to derive in Part C an MEC curve for each possible level of profits in Part A. If the existing profit level is $15, we see in Part B that the desired capital stock will be $340 if $r = 8$ percent, $400 if $r = 6$ percent, and $460 if $r = 4$ percent. Plotting these pairs of values in Part C gives us the MEC curve labeled $Z = \$15$. The MEC curve for $Z = \$40$ is derived in the same way, and this could be repeated for any other level of profits in Part A.

Now suppose that profits are actually $15, the market rate of interest 6 percent, and the actual capital stock $400. We see in Part C that the desired capital stock is $400, so the actual is equal to the desired.[5] Net investment will be zero. As was explained in detail in the preceding chapter, we know in this case that the MEI curve in Part D will be positioned like the curve labeled

[5]It may be objected that this is a highly implausible supposition because profits of $15 on a capital stock of $400 provide a return to stockholders of less than 4 percent, or less than the market interest rate of 6 percent. The answer is that we cannot say anything here about the return to stockholders on the basis of the information given. A major consideration is the amount of capital stock that is debt-financed. For example, if it is $300 of the $400, the profits figure of $15 (which is arrived at after deducting interest paid on debt as well as all other expenses) turns out to provide a 15 percent return on the capital stock that is equity-financed.

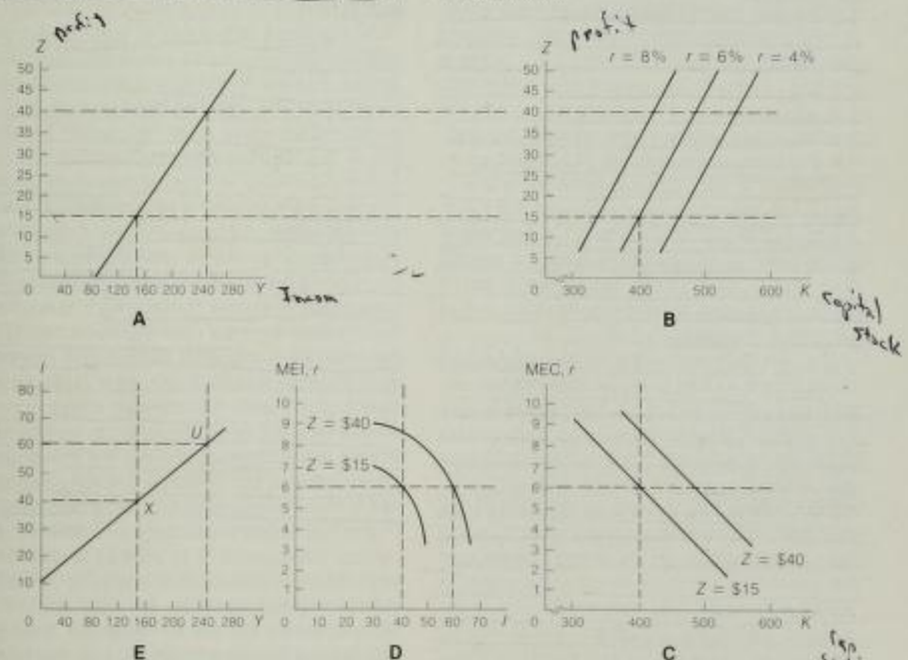


FIGURE 11-1
The level of profits, the desired capital stock, and the investment function

Z = \$15. That curve shows, at the existing interest rate of 6 percent, net investment of zero and replacement investment of \$40, assuming that one-tenth of the existing capital stock of \$400 requires replacement during the period. Finally, noting in Part A that Z of \$15 is found with the economy operating at the \$150 level of Y, it follows that the I of \$40 in Part D is also found with Y of \$150. This combination of I = \$40 and Y = \$150 identifies the point marked X in Part E and establishes one point on the investment function toward which we are working.

By finding other such combinations of I and Y, the investment function in Part E can be fully specified. If we start with the profit level of \$40 in Part A, we will find in Part C that at a 6 percent interest rate the desired capital stock is \$480. With the existing capital stock again assumed to be \$400, the desired exceeds the actual and the MEC exceeds the market rate of interest. The MEI curve in Part D shifts upward to the position shown by the curve labeled Z = \$40; this indicates net investment of \$20 and gross investment of \$60. Since Z of \$40 is found in Part A with the economy operating at Y of \$250, it follows that I of \$60 in Part D is also found with Y of \$250. This identifies the point marked U in Part E. Connecting points like X and U yields the upward-sloping investment curve or function there shown.

The position and slope of the investment curve shown in Part E follow from the specific assumptions made in other parts of the model. With different assumptions, the position and/or slope will be different. For example, if Z made up a larger fraction of Y at each level of Y—i.e., if the curve in Part A were located above the one now shown—the I curve in Part E would then be located above its present position. If the desired capital stock for any given level of profits were larger than now shown—i.e., if the curve in Part B for any given rate of interest were located to the right of its present position—again the I curve in Part E would be located above its present position. If all functions were as shown but the market rate of interest were lower than the 6 percent assumed above, this also would mean that the I curve would be located above its present position. If the curve in Part A were steeper, indicating that the fraction of any change in income that goes into profits is larger than that shown by the present curve, the I curve in Part E would be steeper than the one now shown. If the MEI curve in Part D were flatter (the supply curve of capital goods more elastic), indicating a greater rise in I for any given shift in an MEC curve, again the I curve in Part E would be steeper than the one now shown. However, for present purposes we need not pursue these aspects—it is sufficient to see that the profits theory produces an investment curve of the general type shown in Part E.

One property of this type of curve is an intercept with the vertical axis above zero, but it may appear that the investment curve should have a negative intercept with the vertical axis just as the profits curve does. The profits curve is drawn this way to reflect the fact that at a sufficiently low level of income, aggregate profits will become negative, a rare situation but one that did occur during the Great Depression. A depression so severe that aggregate profits become negative will likely mean negative net investment, but gross investment (at least gross fixed investment) will even then be some positive amount. Some firms will still find it necessary to undertake at least some replacement investment merely to maintain operations at even a very low level. Because the vertical axis of Part E shows gross and not net investment, an investment curve with a positive intercept like the one shown seems to be appropriate.

The amount of investment indicated by the vertical intercept, \$10 in Part E, is ordinarily described as *autonomous*, meaning independent of the level of income. Although in practice income could not fall to zero, the intercept shows what investment would be in this hypothetical case. Then starting at zero and moving to successively higher levels of income, we find

that total investment becomes successively greater. The difference between autonomous investment and total investment is ordinarily called *induced* investment, meaning investment that is called forth by or dependent on the level of income. Thus, in Part E, at Y of \$150, total investment is \$40 composed of induced investment of \$30 and autonomous investment of \$10, and at Y of \$250, total investment of \$60 is composed of \$50 of induced investment and the unchanged \$10 of autonomous investment. According to this terminology, a shift in the investment curve with no change in its slope involves a change in autonomous investment only, and a change in the slope of the curve with no change in its intercept involves a change in induced investment only.

The investment curve we have derived in Part E of Figure 11-1 is duplicated by the dashed curve in Part B of Figure 11-2. (The difference in appearance is due to the fact that the scales of Part B of Figure 11-2 are not the same as those of Part E of Figure 11-1.) Autonomous investment of \$10 is identified by the curve labeled I_a. Induced investment is shown in Part B by the difference between the line I_a and the dashed curve labeled $I_a + eY$. The same separation of investment into autonomous and induced may be seen in Part A of Figure 11-2. Here we find the sum of investment and consumption expenditures measured vertically, whereas in Part B we find only investment expenditures. Since autonomous investment is \$10 at all levels of income, the line in Part A labeled $C + I_a$ lies \$10 above the line labeled C at all levels of income; since induced investment varies with the level of income, the difference between the line labeled $C + I_a$ and the line labeled $C + I_a + eY$ varies directly with the level of income.

Note that the investment function we are now describing has the same properties as the short-run consumption function developed in Chapter 4. The general equation for the short-run linear consumption function was given there as $C = C_a + cY$, where C_a represented autonomous consumption, or the amount of consumption spending that is independent of the level of income, and cY represented induced consumption, or the amount of consumption spending that depends on the level of income. Similarly, the general equation for an investment function of the type given by the curve in Part E of Figure 11-1 or by the dashed curve of Part B of Figure 11-2 is $I = I_a + eY$, in which I_a represents autonomous investment, or the amount of investment spending that is independent of the level of income, and eY represents induced investment, or the amount of investment spending that depends on the level of income.

In the consumption function, induced consumption equals cY, where c indicates the marginal propensity to consume, or $\Delta C/\Delta Y$. In the investment function, we have in e an analogous concept known as the marginal propensity to invest, MPI, or $\Delta I/\Delta Y$.[6]

The Equilibrium Level of Income and Output with Induced Investment

The level of income and output that provides equilibrium is that level at which aggregate spending equals aggregate output. This is the principle elaborated in Chapters 4 and 5 for the two-sector model. In that model we had no theory of investment spending—we simply took investment to be entirely autonomous with no explanation of what might cause changes in the amount of investment. At this point we have a theory of investment spending—that investment depends on the level of aggregate profits—and we can now incorporate that theory into the model of income determination for the two-sector economy.

In Chapters 4 and 5, our model amounted to the following:

[6]Since c has been used for the MPC and s for the MPS, this mnemonic approach suggests i for the MPI. But since this letter has already been used with another meaning, e will be used instead.

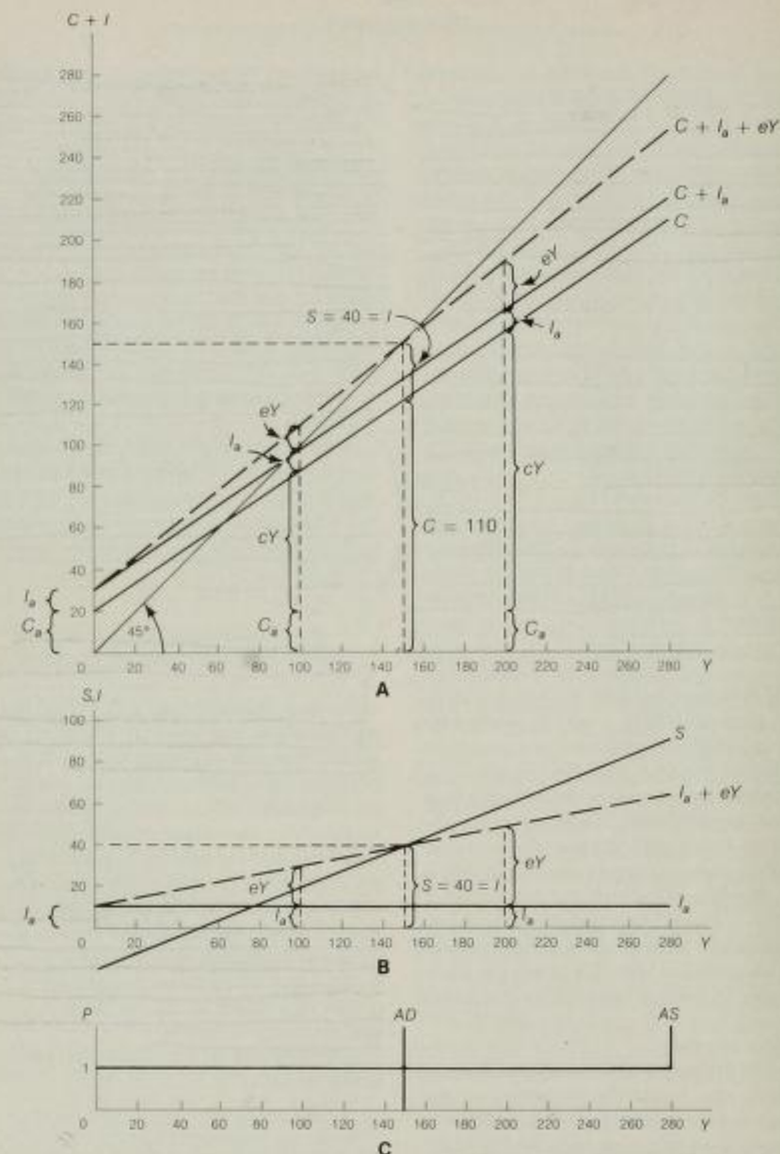


FIGURE 11-2
The equilibrium level of income with induced investment

$$Y = C + I$$
$$C = C_a + cY$$
$$I = I_a$$

so that

$$\boxed{Y = C_a + cY + I_a}$$

With an investment function that makes total investment depend in part on the level of aggregate profits and therefore on the level of the economy's income, we have the following:

$$Y = C + I$$
$$C = C_a + cY$$
$$I = I_a + eY$$

so that

$$\boxed{Y = C_a + cY + I_a + eY}$$

The consumption function in Figure 11-2 is $C = 20 + \frac{3}{5}Y$, and the new investment function is $I = 10 + \frac{1}{5}Y$. Therefore, the equilibrium level of income is readily determined as follows:

$$Y = 20 + \tfrac{3}{5}Y + 10 + \tfrac{1}{5}Y$$
$$Y - \tfrac{3}{5}Y - \tfrac{1}{5}Y = 30$$
$$\tfrac{1}{5}Y = 30$$
$$Y = 150$$

As shown in Part A of Figure 11-2, with Y of 150, C is 110 and I is 40. When output is 150, aggregate income is 150; and, when aggregate income is 150, consumer spending is $20 + \frac{3}{5}(150)$, or 110, and investment spending is $10 + \frac{1}{5}(150)$, or 40. Aggregate spending of 110 + 40, or 150, exactly equals aggregate income of 150. This equality is one way of defining the equilibrium level of income and output.

We may also determine the equilibrium level of income and output by finding that level which equates planned saving and planned investment.

$$S = -20 + \tfrac{2}{5}Y$$
$$I = 10 + \tfrac{1}{5}Y$$
$$-20 + \tfrac{2}{5}Y = 10 + \tfrac{1}{5}Y$$
$$\tfrac{1}{5}Y = 30$$
$$Y = 150$$

Any other level of income and output would be a disequilibrium level. An assumed level below \$150 would mean a lower level of profits and, via the profits theory, a lower level of profits would generate smaller investment. However, as may be confirmed by calculation, investment and consumption would still add up to a larger total than the assumed lower level of output; or, expressed in terms of saving and investment, investment would exceed the amount of planned saving at that assumed lower level of output. This disequilibrium would lead to an expansion of output and income. An assumed income and output level above \$150 would be the opposite kind of disequilibrium and would result in a contraction of income and output. The income level of \$150, and only this level, is the one at which the flow of profits is just the flow needed to generate investment of \$40, which when added to consumption of \$110 produces total spending equal to output of \$150 and thus produces equilibrium. Only at this income level is the flow of profits just the flow needed to generate \$40 of investment or an amount equal to the \$40 of saving at this level of income.

Retaining the assumptions of Part Two, the aggregate demand curve in Part C of Figure 11-2 is shown to be perfectly inelastic and the aggregate supply curve is shown to be perfectly elastic up to the capacity level of output. Given the equilibrium of 150 found in Part A in terms of aggregate spending and found in Part B in terms of planned saving and planned investment, a perfectly inelastic AD curve is established in Part C at Y of 150. With the AS curve perfectly elastic at P of 1 up to Y of 280, Part C shows an intersection between AD and AS at P of 1 and Y of 150. Recall that, on the assumptions that here underlie the AD and AS curves, the equilibrium level of Y is in effect determined in Part A or Part B and simply carried down to Part C.

Changes in the Equilibrium Level of Income

To the extent that investment spending rests on aggregate profits and to the extent that aggregate profits are tightly linked to the

level of income, we cannot explain changes in income as a result of changes in investment induced by changes in aggregate profits. Income changes initiated by changes in investment could then originate only with changes in autonomous investment. In terms of the investment equation, $I = I_a + eY$, a change in income initiated by a change in investment would have to come from a change in I_a or, graphically, through a shift in the investment function. This is the same proposition noted in Chapter 5 for the consumption function. Given the equation $C = C_a + cY$, a change in income initiated by a change in consumption would have to come from a change in C_a or, graphically, through a shift in the consumption function.

However, once a change in income is initiated by a change in either I_a or C_a, the profits theory will come into play. An initial change in income will give rise to induced investment via the marginal propensity to invest. It will also give rise to induced consumption via the marginal propensity to consume. Induced investment and induced consumption produce further increases in income, and these increases in income induce still more investment and consumption spending, a process that continues until a new equilibrium is established. Since the original equilibrium level was derived from the aggregate spending function, $Y = C_a + cY + I_a + eY$, any change in the equilibrium level of income—that is, ΔY—must equal the sum of the changes in the components of the aggregate spending function. That is,

$$\Delta Y = \Delta C_a + c\,\Delta Y + \Delta I_a + e\,\Delta Y$$

This equation may be manipulated as follows:

$$\Delta Y - c\,\Delta Y - e\,\Delta Y = \Delta C_a + \Delta I_a$$

$$\Delta Y(1 - c - e) = \Delta C_a + \Delta I_a$$

$$\boxed{\Delta Y = \frac{1}{1 - c - e}(\Delta C_a + \Delta I_a)}$$

Assuming a rise in autonomous investment from 10 to 20, or ΔI_a of 10, and retaining the values adopted earlier for the MPC and MPI, we have

$$\Delta Y = \frac{1}{1 - 3/5 - 1/5}(0 + 10)$$

$$\Delta Y = 50$$

The previous equilibrium level was 150. With ΔY of 50, the new level is 200. An increase of 10 in autonomous investment has raised the equilibrium level of income by 50.

The results are shown in Figure 11-3. The upward shift in the aggregate spending curve of Part A from $C + I_a + eY$ to $C + I_a + \Delta I_a + eY$ and the upward shift in the investment curve of Part B from $I_a + eY$ to $I_a + \Delta I_a + eY$ raises the equilibrium level of income from 150 to 200. In Part C this produces a shift of the *AD* curve from AD_I to $AD_{I+\Delta I}$. The initial equilibrium at 150 given by the intersection of AD_I and *AS* is replaced by the new equilibrium at 200 given by the intersection of $AD_{I+\Delta I}$ and *AS*.

The introduction of induced investment, eY, into the model has modified the multiplier expression from $1/(1 - c)$ to $1/(1 - c - e)$, as shown above. Since the MPI is positive, the multiplier is larger than it would have been without induced investment. If investment did not respond to a rise in income, the multiplier would have been $1/(1 - 3/5)$, or 2.5, and the rise in income resulting from a rise in autonomous investment of 10 would have been only 25.

The multiplier that emerges from a model in which there is both induced consumption and induced investment has sometimes been called the super-multiplier. As noted, it is necessarily greater than the simple multiplier, given a positive MPI, for if an initial rise in income, produced in our example by a rise in autonomous investment, leads not only to induced consumption spending but also to induced investment spending, the overall rise in income will be greater than if only consumption spending had responded to the rise in income.

Thus, in our example, instead of only induced consumption of 0.6 for every increase in income of 1, we now have induced consumption of 0.6 *plus* induced investment of 0.2, or a total of induced spending of 0.8, for every increase in income of 1. The MPC of 3/5 and the MPI of 1/5 may be combined to give what may be called the "marginal propensity to spend," here equal to 4/5. The super-multiplier may then be

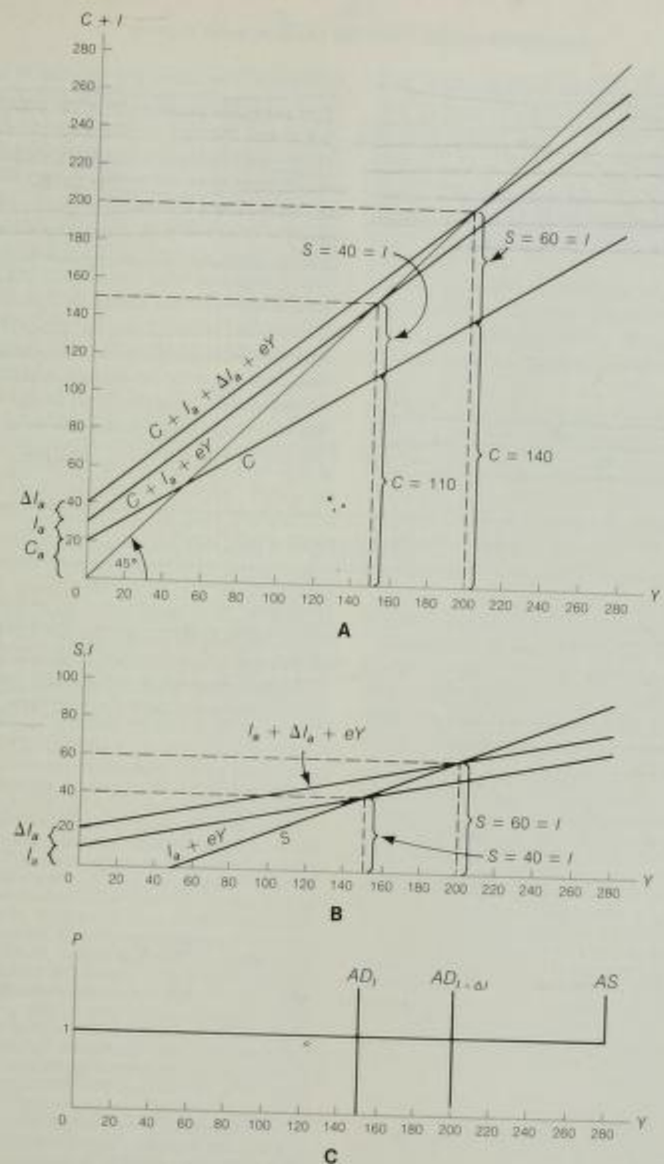


FIGURE 11-3

Effect of a shift in the investment function on the equilibrium level of income: the super-multiplier

alternatively expressed as the reciprocal of 1 minus the marginal propensity to spend, and in our example this is 5. This means that to achieve equilibrium, with an MPC of 3/5 and an MPI of 1/5, or with a marginal propensity to spend of 4/5, income must rise by five times the increase in autonomous investment.[7] Larger or smaller values for the MPI, given the MPC, will, of course, produce larger or smaller values for the super-multiplier and therefore larger or smaller changes in the equilibrium level of income for any given change in autonomous demand.

The profits theory of investment on which we have built the present model of income determination suggests that the super-multiplier will be larger, the larger the change in aggregate profits for any given change in the level of income and the larger the change in investment for any given change in aggregate profits. That is, according to the profits theory, the MPI or $\Delta I/\Delta Y$ is equal to $(\Delta Z/\Delta Y) \cdot (\Delta I/\Delta Z)$ and will therefore vary directly with the size of each of these ratios. Earlier in this chapter we gave some reasons why investment might be expected to vary with profits and noted the way that profits ordinarily vary with income. We have just seen how, given certain assumptions, these relationships may be simply expressed in an investment equation that may be combined with the standard consumption equation to give us the model that was examined here.

THE ACCELERATOR THEORY[8]

The rationale of the profits theory is that the behavior of realized profits guides business persons in judging what future profits will be. In the absence of clear-cut evidence to the contrary, higher current profits carry the promise of higher future profits; they therefore provide business persons with the incentive (and also all or some of the funds needed) to acquire the additional capital goods that will provide the greater output from whose sale those greater future profits will be generated. The accelerator

[7] As in our previous models, the new equilibrium is reached at that income level at which planned saving equals planned investment. But, with induced investment supplementing autonomous investment, the rise in income must be greater to call forth additional saving equal to the sum of the increase in autonomous and induced investment. Thus, in our example, it is only after income has risen by 50 that planned saving will have been brought once again into equality with planned investment. With ΔY of 50, ΔI is 20, made up of the rise in autonomous investment of 10 and in induced investment of $\frac{1}{5}(50)$, or 10. With ΔY of 50, ΔS is $\frac{2}{5}(50)$, or 20. Any smaller increase in income will show an excess of planned investment over planned saving and so a disequilibrium that leads to a further rise in income. Assuming ΔY of 40, we have ΔI of 18, made up of the increase of 10 in autonomous investment and of $\frac{1}{5}(40)$, or 8, in induced investment. With ΔY of 40, we have ΔS of $\frac{2}{5}(40)$, or 16. Therefore, $\Delta I > \Delta S$ $(18 > 16)$, and income expands further. In terms of aggregate spending and aggregate output, with ΔY of 40, aggregate output is 190, and aggregate spending is 192, made up of $C = 20 + \frac{3}{5}(190)$, or 134, plus $I = 20 + \frac{1}{5}(190)$, or 58. With aggregate spending exceeding aggregate output, output and income expand further. With the increase of 10 in autonomous investment, aggregate spending will equal aggregate output, or planned saving will equal planned investment only when income and output increase by 50.

[8] There is a large body of literature on the accelerator theory or the acceleration principle, some of which goes back to the early years of this century. The best known of the early studies is J. M. Clark, "Business Acceleration and the Law of Demand," in *Journal of Political Economy*, March 1917, pp. 217–35. Among the recent studies are R. Eisner, "Capital Expenditures, Profits and the Acceleration Principle," in *Models of Income Determination, Studies in Income and Wealth*, Volume 28, Princeton University Press, 1964; E. Kuh, *Capital Stock Growth: A Microeconomic Approach*, North Holland, 1963; and B. Hickman, *Investment Demand and U.S. Economic Growth*, Brookings Institution, 1965. For a coverage of the pre-1952 literature as well as an examination of the principle itself, see A. D. Knox, "The Acceleration Principle and the Theory of Investment: A Survey," in *Economica*, Aug. 1952, pp. 269–97. Considerable econometric work has been done on the accelerator theory in more recent years; references to this work will be found in D. W. Jorgenson, "Econometric Studies of Investment Behavior: A Survey," in *Journal of Economic Literature*, Dec. 1971, pp. 1111–47.

theory to which we now turn denies that the level of profits plays this strategic role. Its rationale is that the incentive to acquire more capital goods arises not because the current profit record is favorable but because increases in output are putting pressure on firms' existing productive capacity. An increase in productive capacity requires an expansion of the capital stock, which in turn calls for a higher rate of investment spending than would otherwise be needed. In this way, the rate of investment spending is made to depend on changes in the level of output.

When one looks at industry and company statistics, he finds that profits and investment spending are directly correlated. Industries and firms with greater profits spend more on investment than those with smaller profits. There are persons who would conclude from this that the greater investment spending is *due to* the greater profits. The supporters of the accelerator theory see it quite differently. Those industries and companies with relatively high current profits are also those whose sales and output have been expanding relatively rapidly. They then acquire additional plant and equipment not because the level of profits is what it is but because the growth in sales and output is what it is.

From this point of view, investment is explained largely on a physical or input–output basis; increases in the desired capital stock occur because a growing demand for output necessitates a growing supply of the services of capital goods. If the existing capital goods are already utilized, a growing supply of the services provided by such goods can be secured only through an expansion in the stock of these goods. Once again, this brings us back to the accelerator theory argument that the rate of investment spending is tied to changes in the level of output.

The dispute between those who find profits to be the explanation and those who find the changes in output and degrees of capacity utilization to be the explanation, as well as between these and others who offer still different explanations, can only begin to be resolved through econometric study. For example, some light on the particular dispute here noted is shed through the following fairly crude approach. Given the available data on profits, sales, output, and plant and equipment expenditures of thousands of firms, one may compare firms whose changes in sales and output are quite similar but whose profit records are quite different. To the extent that the investment behavior of this group of firms is quite similar, there is some indication that changes in output are more important than the level of profits as a determinant of investment. On the other hand, if one were to compare a group of firms similar in terms of profits but dissimilar in terms of sales and output, a finding of similarity in investment behavior would suggest that profits are more important than changes in output.

While this approach would help separate the effect of the two factors in question, it does not take into account other factors that may be correlated with one or the other of these two and that may actually be the major factors in any explanation. For this purpose, sophisticated statistical techniques are required. In recent years, with the aid of modern computers, these techniques have been applied to masses of time series and cross-sectional data on firms and industries, but even so the complications and difficulties are so great that findings are still not beyond dispute.[9] However, the findings are clear enough to lead most students in this area to the conclusion that the accelerator theory in refined form is one of the most worthwhile theories of investment spending. Because of the difficulties involved, in what follows we will not get into the econometric testing of the theory or into the theory itself beyond its basic form.

[9] For an impression of what is being done in this technical area, see D. W. Jorgenson, *op. cit.* For an essay on the extraordinary difficulties faced by those working in this area, see R. Eisner, "Investment and the Frustration of Econometricians," in *American Economic Review*, May, 1969, pp. 50–64.

Even in this form it will be seen that the fairly straightforward statement that investment spending depends on changes in the level of output is not without its share of complications.

The Capital–Output Ratio

According to the accelerator theory, investment occurs to enlarge the stock of capital because more capital is needed to produce more output. Within limits, firms may be able to produce more output with existing capital through more intensive use of that capital, but there is at any time a particular ratio of capital to output that firms consider optimum. This ratio will vary considerably from industry to industry—much more capital is used per dollar of output in the automobile industry than in the barbering industry—but at any time we will find a particular ratio that is the desired ratio for the economy as a whole. Over time this ratio will change as the mix of output changes—more cars and fewer haircuts or vice versa—and as technological changes and changes in relative factor costs alter the least-cost combination of labor and capital used in the production of different kinds of goods and services. But to reduce the complications, we will assume that this ratio remains unchanged over time.

With K representing the capital stock, Y the level of output (or the physical amount of goods and services produced), and w the capital–output ratio (or the desired number of dollars of capital per dollar of output per time period), we have

$$K = wY$$

If the capital–output ratio is 2, then K of \$400 is desired for Y of \$200, and K of \$450 for Y of \$225.[10] Because we are assuming an unchanging capital–output ratio over time, the desired stock of capital will change over successive time periods *only* with changes in output. Designating some particular time period as t, preceding time periods as $t-1$ and $t-2$, and subsequent time periods as $t+1$ and $t+2$, let us suppose that in period $t-1$ precisely the desired capital stock was on hand to produce the level of output of period $t-1$. That is,

$$K_{t-1} = wY_{t-1}$$

If output then rises from Y_{t-1} to Y_t, the desired capital stock would also rise, from K_{t-1} to K_t, or

$$K_t = wY_t$$

The increase in the desired stock is $K_t - K_{t-1}$. To increase the capital stock, net investment expenditures are needed. To increase the capital stock during t from K_{t-1} to K_t, the net investment expenditures required equal the change in capital stock, or

$$I_t = K_t - K_{t-1} \qquad [1]$$

in which I_t is *net* investment for period t. By substituting wY_t for K_t and wY_{t-1} for K_{t-1}, we may also describe net investment expenditures required in period t as

$$I_t = wY_t - wY_{t-1} = w(Y_t - Y_{t-1}) \qquad [2]$$

This equation simply says that net investment during t depends on the change in output from $t-1$ to t multiplied by the capital–output ratio, w.[11] If $Y_t > Y_{t-1}$, the equation indicates that

[10] Note that K is a stock and Y is a flow. Because flows can be expressed in terms of different time periods, the capital–output ratio will vary with the time period employed. Thus, if it takes \$400 of K to produce \$200 of Y over the course of a year, it will also take \$400 of K to produce \$50 of Y over the course of a quarter of a year (assuming production is at a constant rate over the year). The former indicates a capital–output ratio of 2 and the latter a ratio of 8, but these are seen to amount to the same thing when allowance is made for the difference in the time periods over which output is being measured.

[11] It may be noted that Equations [1] and [2] represent our first encounter with so-called difference equations. By assuming that our variables have only a discrete set of possible values and that these values are available at certain uniformly spaced time intervals, we may date all the variables in our equations. Equations with variables dated in this fashion are one type of difference equations. A set of such difference equations makes up a dynamic model—"dynamic" in that the value of a variable, say, for the time period t, is made dependent on the values of one or more other variables for time periods $t-1$, $t-2$, and so forth. For a discussion of the use of difference equations in economics, see W. J. Baumol, *Economic Dynamics*, 3rd ed., Macmillan, 1970, Part 4.

there is positive net investment during period t; if $Y_t < Y_{t-1}$, there is negative net investment, or disinvestment, during period t. In short, for any assigned value of w and for any change in Y from $t - 1$ to t, the equation indicates the amount of net investment, positive or negative, for period t that is attributable to the change in the level of output.[12]

If we wish to show gross rather than net investment, all we need do is add replacement investment to both sides of the equation. Taking replacement investment to be equal to depreciation and letting D_t represent depreciation in period t, we have

$$I_t + D_t = w(Y_t - Y_{t-1}) + D_t$$

Because negative net investment in plant and equipment is limited to the amount of depreciation of the capital stock, the sum of I_t and D_t cannot be less than zero. If I_{gt} represents gross investment in period t, we may also write

$$I_{gt} = w(Y_t - Y_{t-1}) + D_t \qquad [3]$$

Investment may be expected to respond to changes in the level of output in the way indicated by the equation only if certain assumptions are satisfied. Others will be noted later, but one of the crucial assumptions is best introduced here. This is the absence of excess capacity. We may allow for this factor by letting X_t stand for excess capacity at the beginning of period t and rewriting Equation [3] to read

$$I_{gt} = w(Y_t - Y_{t-1}) + D_t - X_t$$

Whatever the level of gross investment might otherwise be in period t, it will be reduced by the amount of X_t. If the value of $w(Y_t - Y_{t-1}) + D_t$ happened to be equal to or less than X_t, I_{gt} would be zero, the minimum possible figure for gross investment in plant and equipment.

These relationships, which we have here little more than listed in terms of equations, will be discussed more fully and traced through with the help of numerical examples in the following sections.

The Acceleration Principle

The basic relationship with which we are concerned—that between the change in the level of output and the volume of investment spending—is known as the *acceleration principle*. The capital–output ratio, w, is known as the *accelerator*. The theory of investment based on this relationship is known as the *accelerator theory*.

The acceleration principle as expressed in equation form in the preceding section is straightforward. If the economy is already producing the most that can be produced with the existing capital stock (that is, there is no excess capacity or $X_t = 0$), and if there is a fixed ratio between output and capital (that is, w is a constant), it is not hard to see that any expansion of output requires an expansion of the capital stock. Furthermore, if the accelerator has a value greater than 1, the needed increase in capital stock must exceed the increase in output, so that the increase in investment spending will be greater than the increase in output that

[12] This formulation may be critized because it requires that net investment or the change in the stock of capital goods occur instantly, whereas in practice this is a production process that requires time. The problem is this. From Equation [2], we do not know what I_t will be until we know what the change from Y_{t-1} to Y_t is, and we cannot know this before the end of period t when we have the figure for Y_t. This hardly leaves time during period t to carry out the investment, I_t, that is indicated for that time period, and thus constitutes a valid objection to Equation [2]. The problem may be met in one way by making the desired capital stock in any period depend on the output of the preceding period, or $K_t = w(Y_{t-1})$ and $K_{t-1} = w(Y_{t-2})$, from which we derive the investment equation, $I_t = w(Y_{t-1} - Y_{t-2})$, which is not subject to the objection noted. However, this lag system introduces other problems that are better avoided here. Another way of meeting the problem is to make the output of any period depend on the sales of the preceding period. In simplest form, business in each time period chooses to produce output equal to sales of the preceding period, or $Y_t = \text{Sales}_{t-1}$. (This also introduces problems of its own, but these will not bother us here.) Therefore, at the beginning of period t, business knows how much it wants to produce during that period or it knows what Y_t will be. It therefore also knows at the beginning of period t how much net investment is called for during that period to make the actual capital stock equal to the desired stock, namely, $I_t = w(Y_t - Y_{t-1})$, so that Equation [2] as given above is now acceptable.

causes it. Otherwise expressed, to the extent that the demand for additional plant and equipment is derived from the demand for output, a change in the demand for output, given an accelerator greater than 1, leads to a magnification of the derived demand for the plant and equipment necessary to the production of additional output.

To observe the acceleration principle in operation, let us now trace changes in output and gross investment over a number of time periods. In Table 11-1, column (1) simply indicates a series of time periods, and column (2) gives the assumed level of income and output in each period. The specific period-to-period values in this column have been deliberately selected to bring out certain relationships indicated by the acceleration principle; the table does not at all explain why these values are what they are but merely takes them as given.[13] We assume a constant capital–output ratio, *w*, of 2 so the desired stock of capital given in column (3) is two times each period's output as given in column (2). The table includes conditions under which the actual capital stock will exceed the desired capital stock, so column (4) has been inserted to show the actual stock. The average durability of capital goods is here assumed to be 20 time periods instead of 10 as before, so that in each time period there is replacement investment equal to 5 percent of the capital stock in existence in period 1. This gives us in column (5) an unvarying replacement investment of 20 per time period.[14] Net investment in any period, as shown in column (6), equals *w* times the change in output between that period and the preceding period. Gross investment of

[13]According to footnote 12, the output figure for any period is equal to the sales figure for the preceding period; however, these sales figures are not explained here either but are taken as given.

[14]Replacement investment remains at 20 per time period despite the rise in capital stock in period 3 and subsequent periods. The 20 added to the capital stock in period 3 does not need replacement until period 23, the 20 added in period 4 does not need replacement until period 24, and so forth. These are all time periods beyond those given in the table. Note also that the capital stock of 400 on hand in period 1 must have been built up through net investment of 20 during each of the 20 periods preceding period 1 in order to produce the constant 20 of replacement during each time period covered by the table.

TABLE 11-1
The working of the acceleration principle, w = 2

(1) PERIOD	(2) OUTPUT	(3) DESIRED CAPITAL	(4) ACTUAL CAPITAL	(5) REPLACEMENT INVESTMENT	(6) NET INVESTMENT	(7) GROSS INVESTMENT
1	200	400	400	20	0	20
2	200	400	400	20	0	20
3	210	420	420	20	20	40
4	220	440	440	20	20	40
5	250	500	500	20	60	80
6	270	540	540	20	40	60
7	260	520	520	20	−20	0
8	256	512	512	20	−8	12
9	250	500	500	20	−12	8
10	230	460	480	—	—	0
11	200	400	460	—	—	0
12	190	380	440	—	—	0
13	210	420	420	—	—	0
14	220	440	440	20	20	40

column (7) is the sum of replacement and net investment of columns (5) and (6).

With the demand for output unchanged from period 1 to period 2, firms need simply maintain the existing capital stock of 400. This is done by replacing the 20 that wears out during the period. However, when demand for output increases by 10 in period 3, new capital facilities of 20 are wanted. In terms of the equation in which I measures *net* investment only, we have $I_t = w(Y_t - Y_{t-1})$ or $20 = 2(210 - 200)$. Total expenditures for capital goods accordingly rise from 20 in period 2 to 40 in period 3, made up of 20 of replacement and 20 of net investment. With an accelerator of 2, the increase of 10 in expenditures for final output produces an increase of 20 in expenditures for capital goods. In percentage terms, a 5-percent increase in expenditures for final product calls for a 100-percent increase in expenditures for capital goods. It is this relationship that gives the acceleration principle its name.

From period 3 to period 4, output rises by 10 as it did from period 2 to period 3. This indicates net investment of 20 in period 4 to effect the increase of 20 in desired capital. Net investment in period 4 is thus the same as in period 3, and gross investment in period 4 is 40, as it was in period 3. This brings out one of a number of relationships between changes in output and the level of investment suggested by the acceleration principle. In order for gross investment merely to be maintained at the same higher level after it has been increased (e.g., to remain at 40 in period 4 after increasing from 20 in period 2 to 40 in period 3), output must continue to rise from one period to the next. Gross investment can stand still period after period only if output rises period after period. Or to be more precise, gross investment will remain unchanged from one period to the next if the absolute increase in output remains unchanged from one period to the next.[15]

Under what conditions will gross investment increase from one period to the next? Answer: Expenditures for output must increase by ever larger absolute amounts from one period to the next. Thus, gross investment in period 5 rises above that in period 4 because the absolute increase in output from period 4 to period 5 exceeds that from period 3 to period 4. However, note next that despite the further increase in output from period 5 to period 6, gross investment actually declines. The absolute increase in output from period 4 to period 5 was 30, but the absolute increase from period 5 to period 6 was only 20. This illustrates another relationship that follows from the acceleration principle. A mere decrease in the absolute amount of increase in the level of output will lead to an absolute decrease in the level of gross investment. For gross investment to show any absolute increase period after period, the economy's output must show successively larger absolute increases in output period after period. Roughly speaking, the economy must run faster and faster in order for gross investment spending to move ahead at all.

As a next step, the economy's output is assumed to begin a decline in period 7. Output in that period drops 10 below the output of period 6, which means that desired capital in period 7 is 20 less than in period 6. *Net* investment is therefore -20 in period 7. Because 20 of the capital carried over from period 6 will wear out during period 7, business is able to work down the capital stock to the desired lower level simply by not replacing the 20 that wears out during period 7. Gross investment is zero for the period.

Output continues to decline in period 8, but

[15]This requires qualification, because it is true for only a limited number of periods. In the present illustration, in order for gross investment to remain at 40 from period 3 through, say, period 22, output must rise by 10 in each succeeding period. However, in period 23, gross investment will remain at 40 even though output does not rise above its level (400) reached in period 22. In period 23, the 20 of net investment of period 3 comes up for replacement, and replacement investment jumps abruptly from 20 to 40. Similarly, the 20 of net investment of period 4 comes up for replacement in period 24, to make replacement investment 40 in period 24. And so forth.

we now find that, despite this, gross investment increases. This is the result of the fact that there is a smaller absolute decrease in output in period 8 than there is in period 7. We have here just the reverse of the relationship found in periods 5 and 6, where gross investment decreased because, although output was increasing in both periods, the absolute increase was less in period 6 than in period 5. Viewed on the downside, this relationship suggests that an upturn in gross investment need not necessarily await an upturn in output—it may occur even in the face of a decline in output once that decline begins to proceed more slowly. This, it may be noted, helps to explain a phenomenon observed in some business cycles: the fact that the peaks and troughs in real expenditures for capital goods will occur earlier than the peaks and troughs in real expenditures for final output as a whole.

To bring out another important feature of the acceleration principle, it is next assumed that the slowing of the decline in output is followed by a speedup in the decline. Starting in period 8, the absolute decrease in output in the next three periods is 6, 20, and 30. A decrease in output greater than 10 in any period presents a situation not confronted earlier. For example, the decrease in output from 250 in period 9 to 230 in period 10 reduces desired capital from 500 to 460. However, for the economy as a whole, the maximum amount by which the capital stock can be reduced in any period is the amount of the goods that wear out. Individual firms may be able to cut back more rapidly by selling unwanted capital goods to other firms, but this is plainly not possible for all firms combined. In our illustration, the amount that wears out in each period is 20, so 20 becomes the maximum possible net disinvestment per period. Therefore, in period 10 a discrepancy appears between desired capital and actual capital, the former having declined by 40 from period 9 but the latter by only the maximum possible 20. Another decline in output greater than 10 occurs in period 11; this further enlarges the discrepancy between required and actual capital stock. Finally, the figures assumed in the subsequent periods are such that the discrepancy is fully removed in period 13.

The discrepancy found over periods 10 through 12 means that firms are operating with excess capacity during these periods (or that X_t exceeds $w(Y_t - Y_{t-1}) + D_t$ in each period). Thus, the slowing in the absolute amount of decline in output in period 12 is not sufficient to raise gross investment to a positive figure as it did in period 8 when there was no excess capacity present. Not even an absolute increase in the level of output as in period 13 is sufficient to lift gross investment above the zero level in the face of the existing excess capacity. It is not until the excess capacity is eliminated that the acceleration principle becomes operative once again; this occurs in period 14.

Although the acceleration principle becomes temporarily inoperative during periods of excess capacity, the simple mechanics traced for the other periods of Table 11-1 still show that the principle is capable of explaining the relatively wider fluctuations that occur in the expenditures for capital goods than in the expenditures for final goods in general, a real-world phenomenon that economists have long recognized. However, even in those periods during which the economy is operating with no excess capacity, the results shown by the table can be produced only by making certain other assumptions, some of which may be unrealistic. It is clearly necessary to look at these assumptions in order to evaluate the practical significance of the principle.

Closely related to the assumption that firms are operating without excess capacity is the assumption that firms will increase capacity to meet every increase in real spending. In effect, this means that business persons act as automatons, responding to an increase in the quantity of goods sold by increasing investment spending and to a decrease in the quantity of goods sold by decreasing investment spending. In practice, however, even if their capital facilities

are operating at capacity, business persons will try somehow to squeeze additional output from existing plant and equipment unless and until they are convinced that the observed increase in the quantity of goods sold is likely to be permanent.

Similarly, if and when an expansion of capital facilities appears warranted, the expansion may not be exactly that needed to meet the *current* increase in sales; it will probably be sufficient to meet the increase in sales anticipated over a number of years into the future. Piecemeal expansion of facilities in response to short-run increases in quantity of goods sold may be uneconomical or even, depending on the industry, technologically impossible. (One cannot add one-half of a blast furnace.)

The assumption of a constant capital–output ratio or accelerator, *w*, is necessary to our simple mechanical model of the acceleration principle, but it also is rather unrealistic. Even if firms could and did automatically adjust their capital stock to each change in current sales, the capital–output ratio would not be constant. An increase in sales might be concentrated at one time on the output of industries whose technology calls for high capital–output ratios and at another time on the output of industries with low ratios. This means that, even in the absence of technological changes, the degree to which investment spending responds to any increase in quantity of output sold depends on the distribution of that increase among the goods of different industries whose output is subject to different capital–output ratios.

This point suggests another qualification. When we disaggregate investment by industries, we may find that investment for the economy as a whole increases even without an increase in the quantity of output sold. Thus, a redistribution of a given total of expenditures among available goods may lead through the acceleration principle to more net investment in industries enjoying the increased spending than disinvestment in those suffering the decreased spending—since, at the limit, disinvestment in any industry cannot exceed the rate at which capital facilities are used up.

Another assumption of the simple acceleration principle is that any gap between the amount of capital desired by business persons and the amount they actually have is closed within a single time period. One objection to this is that it may be physically impossible. As we saw earlier, if the desired capital stock falls below the actual by an amount greater than the amount of depreciation for the period, it will require more than the current period to reduce the actual capital stock to the desired level. During that interval, the acceleration principle becomes inoperative. Note now that a related situation may be confronted in the other direction. If the capacity output of the capital goods industries for one time period is less than the sum of that period's replacement investment and the excess of the desired over the actual capital stock, the gap cannot be closed in one time period. For example, if the capacity of these industries is $50 billion per year with replacement investment currently absorbing $20 billion per year, a gap between desired and actual capital greater than $30 billion cannot be filled within the time period of one single year. However, the simple acceleration principle will show net investment of $40 billion for a gap of $40 billion, despite the fact that this rate of net investment may be impossible.

Beyond this, even if the productive capacity of the capital goods industries were always physically sufficient to close any gap in one time period, it does not follow that this would happen. As we saw in Chapter 10, the amount of net investment spending in one time period may be below the amount needed to close an existing gap during that time period, not because the capital goods industries do not have the physical capacity to produce the amount of capital goods needed but because net investment spending itself is restrained below that amount by the effect on the MEI of the rising cost of capital goods as these industries expand output closer to their capacity levels. We may

$K = wY$

illustrate the simple acceleration principle and this particular qualification of it by building on the graphic apparatus developed in Chapter 10.

The Acceleration Principle and the MEC and MEI Schedules

The accelerator theory of investment and the acceleration principle on which it rests make the desired capital stock proportional to the level of output. If the accelerator is 2, as in Part A of Figure 11-4, the desired capital stock is equal to twice the level of output. The curve in Part A also shows for any change in the output level—e.g., a rise from \$200 to \$220—a rise in the desired capital stock equal to twice the output change—e.g., a rise from \$400 to \$440. Because the acceleration principle in its rigid form assumes that any gap between the desired and actual capital stock is filled within a single time period, it follows that net investment in that single time period will be equal to the accelerator times the change in the level of

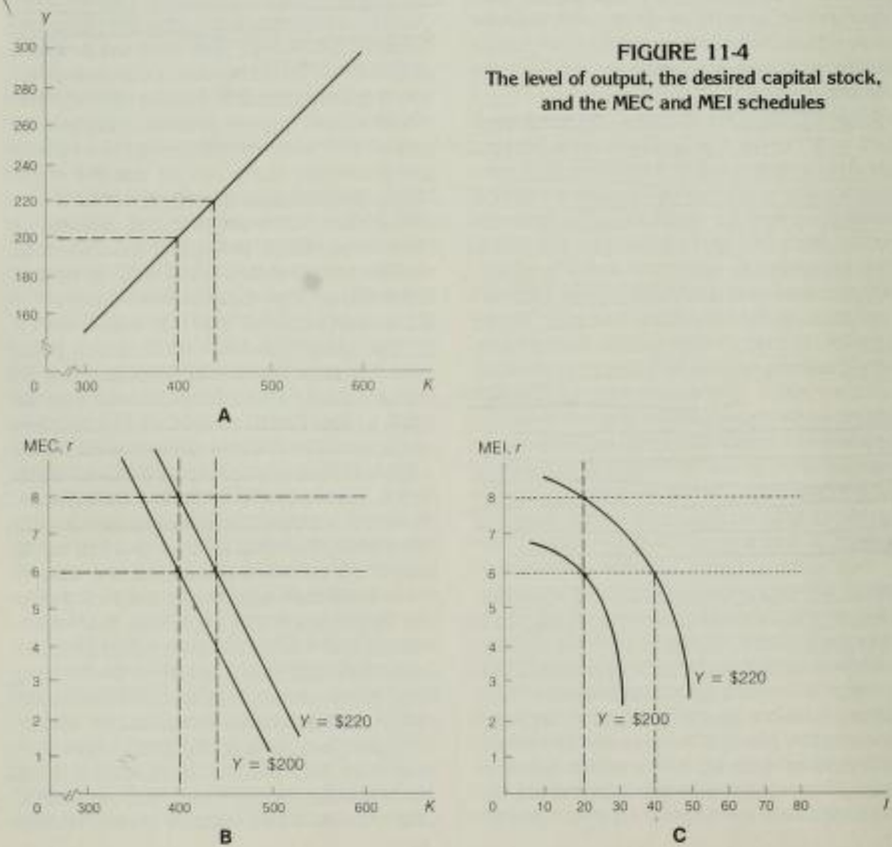


FIGURE 11-4
The level of output, the desired capital stock, and the MEC and MEI schedules

output. In terms of the equation derived earlier, we may write

$$I_t = K_t - K_{t-1} = w(Y_t - Y_{t-1})$$

And in terms of the numerical illustration just given,

$$\$40 = \$440 - \$400 = 2(\$220 - \$200)$$

This result, in which the actual capital stock grows to the desired capital stock within the same period that a gap between the two appears, can occur only if the conditions noted earlier are satisfied. Parts B and C of Figure 11-4 show graphically what is involved.

These two parts bring in the familiar MEC and MEI curves. In Part B, two MEC curves are plotted, one corresponding to Y of \$200 and one to Y of \$220. The acceleration principle alone indicates that the desired capital stock will be \$400 with Y of \$200 and an accelerator of 2. However, we take into account that the desired capital stock will also vary with the rate of interest. With Y of \$200, the curve so labeled shows that desired capital will be \$400 when r is 6 percent, \$420 when r is 5 percent, and \$380 when r is 7 percent. With Y of \$220, desired capital will vary with the rate of interest as shown by the MEC curve labeled $Y = \$220$.

Suppose now that the actual level of Y is \$200, the market rate of interest is 6 percent, and the actual capital stock is \$400. Since the actual and the desired capital stocks are equal, net investment is zero and gross investment is equal to replacement investment. Assuming, as in Table 11-1, that the average life of capital goods is 20 years, replacement investment is \$20. The MEI curve in Part C labeled $Y = \$200$ thus shows gross investment of \$20 and net investment of zero at the market interest rate of 6 percent.

Suppose, next, that the output level rises from \$200 to \$220 so that we move from the lower to the upper MEC curve in Part B. With the interest rate at 6 percent and the actual or existing capital stock still at \$400, desired capital exceeds actual capital by \$40. The MEI curve in Part C labeled $Y = \$220$ shows that in the first time period gross investment will be \$40 and, accordingly, *net* investment will be \$20, or less than the full amount of the desired increase in the capital stock. It will therefore actually take a series of time periods to adjust the capital stock to the desired level.

Since the simple acceleration principle assumes that net investment sufficient to close the gap between actual and desired capital stock will occur in a single time period, no matter how large the amount involved (\$40 in the present illustration), it must also assume, quite unrealistically, that the short-run supply curves of the capital goods industries are perfectly elastic over an unlimited range. In other words, it must assume not only that these firms do not run into short-run rising costs as they expand output but also that their capacity to expand output is unlimited in the short run. This would be described graphically in Part C by the dotted-line MEI curves, which remain perfectly flat over an unlimited range. The assumed rise in output in Part A would then lead to a shift in the MEI curve in Part C from the lower to the upper dotted line. Gross investment, which had been equal to \$20 for the time period in which output was \$200, would now jump to \$60 in the next time period in response to the increase in output to \$220 in that time period. If the MEI curve were actually perfectly elastic, there would be no barrier to carrying out whatever amount of net investment is needed to close the gap between actual and desired capital stock in one time period.

Although the MEI curve may not turn down as soon or as sharply as the solid-line MEI curves illustrated, it remains that it must turn down at some level of investment. Expansion of output by the capital goods industries must sooner or later run into rising marginal costs, given the fact that the productive capacity of these industries is limited in the short run. The assumption of the simple acceleration principle that any increase in the desired capital stock will be met in a single time period is, therefore, one that cannot always be satisfied. The answer to this and other problems that arise from the simple,

rigid version of the acceleration principle is to resort to a more flexible and refined version that can handle them. We will give this matter some attention in Chapter 19 on business cycles. For example, we will see there that the business cycle theory of J. R. Hicks employs the accelerator theory to explain investment spending but not in the unconstrained form we described here. Among other modifications, his accelerator theory recognizes both a limit to the rate of disinvestment during downturns and a limit to the rate of investment during upswings.

Like the profits theory, the accelerator theory, in whatever form it takes, is a theory of investment spending, but, unlike the profits theory, it expresses a dynamic relationship. It makes the level of investment a function of the *change* in another variable, the level of output, and is useful in the kind of dynamic analysis that traces successive changes in the level of output, the subject matter of business cycle theory. In this connection, it may have been noted that we did not come out with a model of the equilibrium level of income in this section on the accelerator theory as we did in the section on the profits theory. In Figure 11-4 we saw how the level of investment spending depends on a change in the level of output, but we could not identify an equilibrium level of real income or output consistent with that particular level of investment because that particular level of investment is what it is because of a *change* in the level of real income and output. What we can and will do in the chapter on business cycles is to see how a change in real income and output that makes one period's investment different from that of the preceding period will in turn affect the real income and output in the next period and possibly set into motion an interaction over time that causes output to rise and fall as it does during business cycles.

The profits theory in the first part of the chapter was also submitted as a theory of investment spending, but it logically led us into a theory of income determination. Because it is a static theory—investment depends on the level of profits and the level of profits depends on the level of income—it fits into a static model of the determination of the equilibrium level of income. We found the equilibrium level of income to be that level at which the sum of the amount of investment called forth by the amount of profits found at that level of income plus the amount of consumption called forth by that income would just add up to that particular figure for income. Thus in Figure 11-1, we had a simple interdependent static system in which we found the values for profits, investment, and the equilibrium level of income.

Although the profits theory and the accelerator theory differ in this and various other ways, they both operate on investment spending by causing shifts in the MEC and MEI curves as has been seen in Figures 11-1 and 11-4. In Chapter 12, we turn to the rate of interest and the role of finance, forces that operate on investment spending by causing movements along *existing* MEC and MEI curves.

chapter 12 13

Investment Spending: the Rate of Interest and the Role of Finance

The extent to which a change in the rate of interest causes a change in the rate of investment spending depends on the elasticity of the MEC schedule. The more elastic the schedule, the greater the increase or decrease in the profit-maximizing capital stock that follows from any decrease or increase in the rate of interest. In order to specify the relationship between a change in the rate of interest and a change in the rate of investment spending, we must proceed from a change in the rate of interest through the resultant change in the profit-maximizing capital stock to the resultant change in the rate of investment.

THE RATE OF INTEREST AND THE RATE OF INVESTMENT

As a first step, it will be helpful to review the relationships found in Chapter 10 between the rate of interest, the MEC schedule, and the MEI schedule. Figure 12-1 parallels Figure 10-2 except for changes in the numerical values and the omission of replacement investment. As in Figure 10-2, with the MEC schedule given in Part A, the actual capital stock given as $400 billion, and the interest rate given as 6 percent, we see that the actual capital stock is also the profit-maximizing capital stock. The rate of net investment as shown in Part B is accordingly zero. Now, if we assume a drop in the rate of interest from 6 to 5 percent, the profit-maximizing capital stock in Part A of Figure 12 becomes $600 billion, and net investment in Part B becomes $5 billion for the first time period following the drop in the rate of interest. Since the addition of $5 billion to the stock of capital during the first period does not move the economy perceptibly down the MEC schedule, the MEI schedule does not shift downward appreciably, and the rate of investment remains virtually the same in the second and third periods and in as many periods as may properly be included in the short run.[1]

[1]It will be recalled that the short run, as defined in Chapter 10, is a time interval in which the growth in the stock of capital is not great enough to depress the MEI schedule appreciably.

Figure 12-2 differs from Figure 12-1 only in the elasticity of its MEC schedule. In Figure 12-2 the MEC declines relatively more rapidly with increases in the stock of capital than in Figure 12-1. Assuming for Figure 12-2 the same initial equilibrium given in Figure 12-1, a drop in the rate of interest from 6 to 5 percent will increase the profit-maximizing capital stock from $400 billion to only $420 billion. Note, however, that, despite the inelasticity of the MEC schedule of Figure 12-2, the rate of investment in the first time period following the drop in the rate of interest is $5 billion in Figure 12-2, just as it is in Figure 12-1. Starting from a position in which the profit-maximizing capital stock equals the actual stock, the *initial* impact of a drop in the rate of interest on the rate of investment spending is independent of the elasticity of the MEC schedule. The reason is that the initial impact of such a drop depends only on the elasticity of the MEI schedule, which is independent of the elasticity of the MEC schedule.[2]

Nonetheless, the elasticity of the MEC schedule must eventually exert an influence on the rate of investment spending. For, if the MEC schedule is relatively inelastic, additions to the stock of capital that result from each period's net investment will move the economy quickly down such a schedule. This we know will push down the MEI schedule and with it the rate of investment at the given market rate of interest. In short, other things being equal, the more inelastic the MEC schedule is, the sooner appreciable downward pressure on the MEI schedule will appear as a "feedback" from the growth in the stock of capital.

Although this analysis shows that the effect of a change in the rate of interest on the rate of investment spending in any time period depends on the elasticity of the MEC schedule, it does not tell us anything about what determines that elasticity. We cannot appraise the significance of changes in the interest rate as one of the factors influencing the rate of investment unless we know something about what determines the elasticity of the MEC schedule.

[2]It will be recalled that the elasticity of the MEI schedule depends on the elasticity of the supply schedule of capital goods. The latter may be assumed to be the same in both cases, so the elasticity of the MEI schedule is the same in both figures. See pp. 165–69.

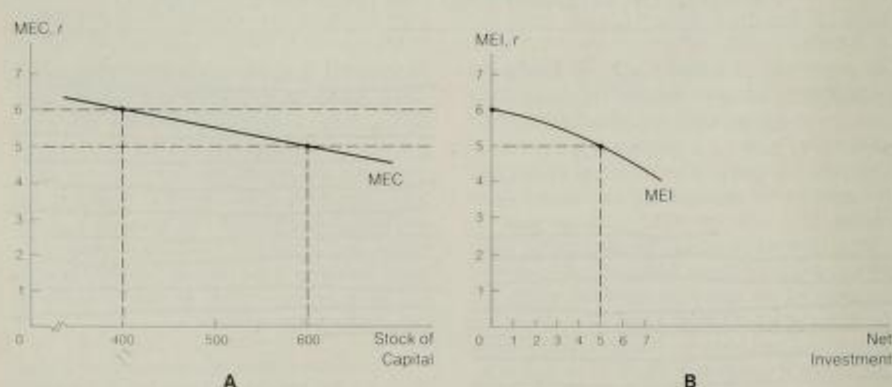


FIGURE 12-1
Net investment with an elastic marginal efficiency of capital schedule

Determinants of the Elasticity of the MEC Schedule

Assuming again that a firm buys capital goods only in the expectation that each purchase will add to its profits, the elasticity of the MEC schedule for the economy as a whole may be said to depend on the significance of any given change in the interest rate as a factor altering the profit-maximizing or desired capital stock for all firms combined. A fall in the interest rate may lead some firms to expand their stock of capital goods substantially, while the same fall in the interest rate may lead other firms to no expansion at all. The elasticity of the aggregate MEC schedule thus depends on the combined responses of all firms to any change in the interest rate.

Our problem, then, is to investigate what determines whether any given change in the interest rate will lead to a larger or smaller change in the profit-maximizing capital stock. Although it is immediately apparent that a lower interest rate, other things being equal, is also a reduction in one cost of doing business, we find that different businesses respond differently to such a drop. In seeking an explanation for this disparate reaction, we will get at the factors that determine the elasticity of the MEC schedule.

With all other things assumed to remain unchanged, especially the existing state of technology, one factor that determines how great an increase in the profit-maximizing capital stock will result from any given reduction in the interest rate is the extent to which firms find it technologically possible to substitute what is now the relatively cheaper capital input for what is now the relatively more expensive labor input. It must be seen here that, in terms of the cost of factor units, the comparison is between the rate of interest as the cost of capital and the wage rate as the cost of labor. Since we express the rate of return from or the productivity of an addition to the stock of capital as a percentage—namely, the MEC—we must for comparability express the cost of capital as another percentage—namely, the market rate of interest. A fall in the rate of interest is then a decrease in the cost of capital. In contrast, since we express the return from or the productivity of an addition of a unit of a given type of labor as the dollar value of the output that will be produced because of that additional unit, the cost of labor is also expressed as a dollar value, the dollar

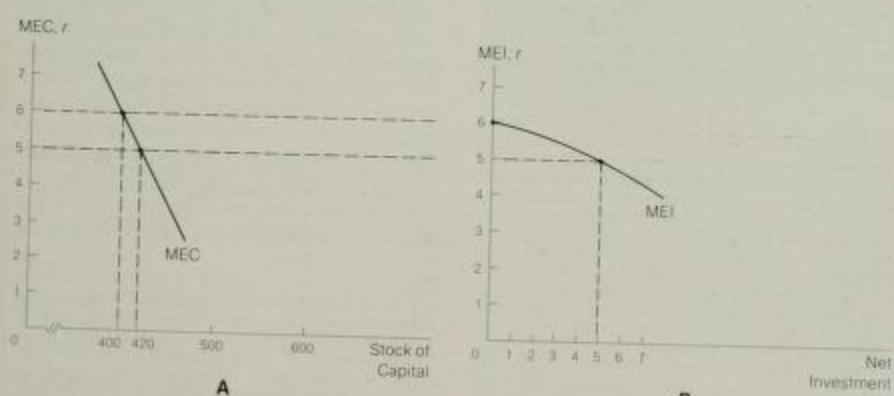


FIGURE 12-2
Net investment with an inelastic marginal efficiency of capital schedule

cost of the services of a unit of the given type of labor. A fall in the wage rate paid labor is then a decrease in the cost of labor.[3]

With a fall in the interest rate and no change in the wage rate, there will be a tendency throughout the economy to substitute the cheaper input, capital, for the more expensive input, labor, in the production of any constant level of output. Although such a tendency will be present throughout the economy, the extent to which substitution of this sort actually takes place will vary widely from industry to industry and, to a lesser extent, from firm to firm within a particular industry.

For some firms the state of technology may be such that, regardless of the fall in the cost of capital, substitution is severely limited or even impossible. At the extreme, for firms whose technology is such that production requires a rigidly fixed combination of men and machines, a fall in the rate of interest will lead to no substitution of capital for labor.[4] For such firms, the amount of capital employed in the production of a given level of output is virtually independent of the interest rate. In this case, as far as substitution goes, a fall in the interest rate will have no effect on the size of the profit-maximizing capital stock consistent with a constant level of output. This also means that the rate of investment spending will be unaffected by a drop in the interest rate.

For other firms, the amount of capital employed may be more or less sensitive to changes in the interest rate. Their technology may be such that, within limits, each drop in the interest rate can make further substitution of capital for labor profitable. To take a crude example that may readily be compared with the one above, suppose that the existing technology permits the firm a choice between a one person–one machine, one person–two machine, or one person–three machine combination. However, to switch to the one person–two machine combination requires purchase of an "Adapter A," which links two of the original machines in such a way that one worker may operate both, and to switch to a one person–three machine combination requires a second "Adapter A" plus one "Adapter B" to link three of the original machines in such a way that one worker may operate all three simultaneously. After the linkage each of the original machines will turn out the same number of units of output as before. Before the fall in the interest rate, the then current costs of capital and labor made the one person–one machine combination the least-cost combination. But with an initial fall in the interest rate, capital becomes relatively cheaper, so one "Adapter A" may be purchased for every two machines; one-half of the labor force employed in operating the machines may be

[3]If we express the productivity of a capital good by its MEC, which is a percentage, we must measure the "price" of that capital good not by the number of dollars paid for the good but by the rate of interest, explicit or implicit, paid for the funds used to purchase that good. The cost element represented by its supply price has already been allowed for in the computation of its MEC. The MEC, it will be recalled, is the percentage rate of return expected on the capital good after allowance for all costs other than interest cost. Once given the MEC, the decision to buy or not to buy the capital good requires a comparison of its price, which is the rate of interest, with its return, which is its MEC.

This, it may be noted, is not the only approach to this question. In contrast to this widely used Keynesian formulation is the neoclassical formulation in which the price of capital is expressed as the price of a unit of capital services, and that in turn is the cost to the firm of using that unit of capital for a period of time. From this approach, the cost of a unit of capital services is expressed as a dollar value, as is the cost of hiring a unit of labor services. See D. W. Jorgenson, "The Theory of Investment Behavior," in R. Ferber, editor, *Determinants of Investment Behavior*, National Bureau of Economic Research, 1967, pp. 129–55.

[4]Even here, however, there is some opportunity for substitution. For example, if a firm has 100 machines and must have 100 workers to operate the machines and if the machines are subject to intermittent breakdown, the fall in the price of capital may lead to the purchase of one or more additional machines to be pressed into use when any of the 100 in use at any time breaks down. With the fall in the price of capital, this increase in the capital stock may be cost saving, for the relatively more expensive input, labor, will not remain idle while the relatively less expensive input, capital, is being repaired. It is true that one or more machines will always be idle or in repair, but this "idleness" may, with the lower interest rate, be less costly than the otherwise idle labor.

replaced by capital. With some further fall in the interest rate, the least-cost combination may call for the purchase of another "Adapter A" and one "Adapter B" for every three original machines; one-third of the remaining labor force employed in operating the machines may be replaced by capital. The firm's output remains the same in all cases.

Our first factor is seen to be essentially technological in nature. With a given state of technology but with some opportunity for firms to vary factor combinations within this given state of technology, one way in which a drop in the interest rate is translated into an increase in the profit-maximizing capital stock is through a substitution of capital for labor in the production of any level of output. The greater the possibilities available within the existing state of technology, the more elastic the MEC schedule will be.

Closely related to this factor is the cost saving represented by a switch toward greater use of capital as a result of a drop in the interest rate. Even though the existing technology may permit variation in the capital–labor combination, the extent to which such variation will occur in response to a change in the interest rate depends on the cost advantage offered by such variation. Two illustrations may be traced through to show in a rough way the operation of this factor. In the first, presented in Table 12-1, a fall in the interest rate leads to a cost advantage through the substitution of capital for labor in the production of a given level of output; in the second, presented in Table 12-2, a fall in the interest rate leads not to reduction in the quantity of labor employed but to the use of more durable capital goods in combination with an unchanged quantity of labor.

In the first illustration, suppose that there are two firms, with outputs of identical value and with identical costs of $1,000 per year. As shown in Table 12-1, Firm A's method of production emphasizes labor, and Firm B's method emphasizes capital. Firm A's method may be described as more labor intensive and less capital intensive than Firm B's, and vice versa. The first part of the illustration assumes a market interest rate of 6 percent, and the second part, one of 4 percent. In both parts, straight-line depreciation with a ten-year life for capital goods is assumed, making annual depreciation cost 10 percent of the value of the capital stock. Annual interest cost is simply the market rate times the capital stock. A drop in the interest rate from 6 to 4 percent reduces Firm A's total costs for the given output from $1,000 to $990, a reduction of 1 percent, and reduces Firm B's total costs for output of the same value from $1,000 to $920, a reduction of 8 percent. Since both firms are assumed to have output per period of equal value, the drop in the interest rate

TABLE 12-1
An illustration of the substitution of capital for labor as a result of a fall in the interest rate

		CAPITAL STOCK	ANNUAL DEPRE-CIATION	ANNUAL INTEREST COST	ANNUAL LABOR AND OTHER COSTS	TOTAL COSTS
Market Rate of Interest 6 Percent	Firm A	$ 500	$ 50	$ 30	$920	$1,000
	Firm B	4,000	400	240	360	1,000
Market Rate of Interest 4 Percent	Firm A	$ 500	$ 50	$ 20	$920	$ 990
	Firm B	4,000	400	160	360	920

produces a 1-percent reduction in cost per unit of Firm A's output and an 8-percent reduction in cost per unit of Firm B's output.

With the interest rate at 6 percent, Firm A's method of production, on a strict cost basis, is as good as Firm B's.[5] But with a drop in the interest rate, Firm B's more capital-intensive method of production has a significant cost advantage over Firm A's more labor-intensive method. Insofar as we can generalize from this illustration for the economy as a whole, the fall in the interest rate will produce some tendency toward greater use of capital and less use of labor in the production of any given level of aggregate output.

In this first illustration, the two firms' methods of production are essentially different: Firm A uses up $50 of capital goods and $920 of labor services in producing its output per year, and Firm B uses up $400 of capital goods and $360 of labor services in producing output of the same value in the same time period. In our second illustration, let us assume two firms whose methods of production are identical in that each uses up the same amount of labor services in the production of output of the same value but whose methods of production differ in terms of the *durability* of the capital goods they employ in combination with the same quantity of labor services.

In Table 12-2, Firm Y has capital stock of $1,000 made up of capital goods with an average life of two years, and Firm Z has capital stock of $3,500 with an average life of ten years. Since both firms produce output of the same value and incur labor costs of the same amount, the capital stocks of the two firms make an equal contribution to output. At first glance, one may object that it is then better to invest $1,000 than $3,500 in capital, if both investments provide capital goods with equal capacity to produce. Is not interest cost at 6 percent on $1,000 only $60, whereas on $3,500 it is $210? This, however, overlooks the fact that the more expensive capital goods, though no more productive per year than the less expensive, are more expensive only because they are more durable and will make the same contribution to production over a greater number of years. Our figures are deliberately selected to show that they are not proportionally more expensive. The purchase price of the more durable capital goods is only 3.5 times that of the less durable, but their life is 5 times that of the less durable. This in turn means that the cost of the more

[5] Considerations other than total cost per unit would, of course, be relevant in choosing the "better" method. For example, the percentage of costs that are fixed is much lower for Firm A than for Firm B. This puts Firm A in a much better position to meet a decrease in demand, since it can reduce total costs almost proportionally with a reduction in output. On the other hand, if wage rates are subject to constant upward pressure, Firm B may be in a better position, since Firm A's costs for any level of output will rise much more sharply. These and other considerations would enter into determining the "better" method.

TABLE 12-2

An illustration of the use of more durable capital as a result of a fall in the interest rate

		CAPITAL STOCK	ANNUAL DEPRE-CIATION	ANNUAL INTEREST COST	ANNUAL LABOR AND OTHER COSTS	TOTAL COSTS
Market Rate of Interest 6 Percent	Firm Y	$1,000	$500	$ 60	$440	$1,000
	Firm Z	3,500	350	210	440	1,000
Market Rate of Interest 4 Percent	Firm Y	$1,000	$500	$ 40	$440	$ 980
	Firm Z	3,500	350	140	440	930

durable capital goods, as measured by annual depreciation, $350, is less than the depreciation cost of the less durable goods, $500, the output from both collections of capital goods being the same for each year.

As in the previous illustration, we may say that, on a strict cost basis, Firm Y's method of production is as good as Firm Z's when the interest rate is 6 percent.[6] Each has output of the same value produced at the same total cost. However, with a drop in the rate of interest from 6 to 4 percent, the second part of the illustration in Table 12-2 shows a decrease in total cost from $1,000 to $980, or 2 percent, for Firm Y and from $1,000 to $930, or 7 percent, for Firm Z. Since the value of output is the same for both firms, Firm Y has a 2-percent and Firm Z a 7-percent reduction in cost per unit of output. With a drop in the interest rate, there is clearly a cost advantage in Firm Z's method, which employs more durable capital goods. Insofar as we can generalize from this illustration for the economy as a whole, the fall in the interest rate will produce some tendency toward production of any given level of aggregate output with more durable capital goods. In other words, there will be a tendency to produce any given level of output with a larger stock of capital goods.

In both illustrations, a fall in the interest rate results in a tendency toward more capital-intensive production. In the first illustration, this was due to a substitution of capital for labor; in the second, it was due to the use of more durable capital goods with no reduction in the quantity of labor used. A fall in the interest rate would, of course, tend to encourage substitution of capital for labor *and* the use of more durable capital goods simultaneously. Both tend to increase the profit-maximizing capital stock consistent with any given level of output, and therefore both work toward a higher rate of investment spending.

[6]This is subject to other considerations of a type noted in footnote 5. Especially relevant would be Firm Z's greater exposure to the dangers of obsolescence.

These, then, are the factors through which a change in the interest rate will affect the profit-maximizing capital stock. Whether the change in the profit-maximizing capital stock will be large enough to produce an MEC schedule that is elastic with respect to any given interest-rate change depends on the combined responses of all firms to that change. It is one thing to indicate the factors that determine the responses of firms and quite another to conclude that the combined responses of all firms for any time period and for any change in the interest rate will produce an aggregate MEC schedule that is elastic or inelastic.

Elasticity of the MEC Schedule—Structures, Producers' Durable Equipment, and Business Inventories

Since the thirties, most economists have argued that the rate of investment spending for an industrially advanced economy such as that of the United States is interest-inelastic. Although this appears to be the consensus, this general conclusion is not equally applicable to all forms of investment spending.

To the extent that investment does respond to changes in the interest rate, the most pronounced response seems to occur in the field of investment spending for residential and business structures. The reasons for this are implicit in the illustrations above. For example, an additional one million square feet of living space or office or factory space per year can be provided only by the construction of the required quantity of houses, apartments, office buildings, or factories.[7] Since the product yielded by construction as such is measured in square feet of floor space available per year, produc-

[7]An existing amount of floor space cannot be substantially "enlarged" by altering factor proportions. More labor applied to an unchanged quantity of floor space can make for more rapid repairs and redecoration and thus for higher utilization of the existing quantity, but possibilities in this direction are very limited.

tion of this product is technologically very capital intensive.[8] The sum of depreciation cost and interest cost per unit of output looms large as a fraction of total cost per unit of output of this kind. Furthermore, production is not only very capital intensive, but it is carried out with capital goods that are durable. To provide another one million square feet of floor space this year requires the construction of buildings that will provide this same output for many years. Since the interest cost on durable capital goods goes on for many years, even small changes in the interest rate can mean a substantial difference in the cost per unit of output (square foot of floor space per year). In other words, where the technology makes for capital-intensive production and where the nature of the capital employed is very durable, a change in the interest rate can make a great difference in cost per unit of output, and this will markedly affect the rate of investment expenditures for this type of capital goods.

The responsiveness of inventory investment to changes in the interest rate is of a different nature. Just as there is a profit-maximizing stock of plant and equipment, there is also a profit-maximizing stock of inventories. However, unlike the case of structures, the profit-maximizing stock of inventories is not likely to be very responsive to interest-rate changes. A lower interest rate reduces the carrying cost of any given amount of inventories, of course, and therefore may be expected to increase the profit-maximizing stock of inventories, other things being equal. However, as we saw above, a fall in the interest rate affects primarily the profit-maximizing capital stock consistent with any level of output by inducing substitutions of capital for labor or by inducing the use of more durable capital equipment. Neither of these factors has any real applicability to the determination of the profit-maximizing stock of inventories, which is determined primarily by the firm's sales, interest cost, and other carrying costs, rather than by its method of production.[9] However, the interest rate will play some role in the determination of the firm's profit-maximizing inventory. Other things being equal, when interest rates rise, firms will attempt to maintain their rate of production and sales with smaller stocks of inventories and a necessarily higher rate of turnover of these smaller stocks, particularly if interest rates rise sharply or to unusually high levels. The response will vary with different firms, depending in part on how large the interest cost of carrying inventories is as a percentage of the firm's total costs for any level of output.

Thus, we may expect the response to changes in the interest rate for the economy as a whole to have the greatest impact on the stock of structures, less on the stock of producers' durable equipment, and least on the stock of business inventories. In view of the fact that different types of investment spending show different responses to changes in interest rates, it is obviously difficult to generalize with respect to the elasticity of investment spending in the aggregate. This is further complicated by other considerations not examined above. Among these is the likelihood of less inelasticity at high levels of the interest rate than at low levels, so that a change in the interest rate from 10 to 9 percent will probably lead to a greater increase in the profit-maximizing capital stock than will an interest-rate change from 5 to 4.5 percent. In other words, whatever the elasticity of the overall schedule, we may expect it to be more inelastic at low interest rates than at high.[10] Its actual elasticity, however, remains an unsettled question.

In summary, we can only repeat the gener-

[8] This is not to be confused with the capital intensity in the production of the buildings that in turn produce the output in the form of a number of square feet of floor space per year. The production method employed in constructing the buildings is an altogether different question.

[9] For a simple model of the determination of the firm's optimal inventory, see W. J. Baumol, *Economic Theory and Operations Analysis*, 4th ed., Prentice-Hall, 1977, Ch. 1.

[10] See A. H. Hansen, *Business Cycles and National Income*, Norton, 1964, pp. 113–38.

ally accepted conclusion reached by Kuh and Meyer some years ago:[11]

> It is difficult to say how sensitive the investment–interest rate relationship is likely to be in the short run, other than presuming it is something greater than zero. Available evidence, none of which is terribly satisfactory, suggests that the interest elasticity of demand is not large, at least in the historically relevant range of roughly 3 to 10 percent *per annum* charged for long-term capital.

THE ROLE OF FINANCE—BEYOND THE INTEREST RATE

So far in this chapter, the role of finance in the investment decision has been drastically simplified by assuming that it involves only one element, the market rate of interest. Investment spending in any time period has been determined from the MEI schedule by drawing a horizontal line at the prevailing market rate of interest to an intersection with the MEI schedule. With a given MEI schedule, only a change in the interest rate would produce a change in investment demand.

Without denying the underlying correctness of this basic relationship, we must recognize the oversimplification that it involves. By showing that aggregate investment spending is determined by the intersection of the MEI curve with a horizontal line drawn at the prevailing interest rate, we are assuming for one thing that funds are available in unlimited amounts to borrowers of every description at this one particular interest rate that we call the market rate of interest. A first objection to this is the fact that, in reality, there is no such thing as *the* market rate of interest. Rather there is a whole complex of rates. The rate paid by any particular borrower depends on such variables as the term of the loan, the size of the loan, the collateral offered, and especially the credit worthiness of the borrower. Beyond this, there is not an unlimited supply of funds available to any single borrower in any time period at this rate; there is some tendency for the rate he pays to rise with each addition to his indebtedness per time period, other things being equal.

A more fundamental objection to this formulation is its implication that the only way in which firms may secure funds to finance investment spending is by borrowing in the market and paying whatever interest rate they must pay. Actually there are three sources of funds for investment, only one of which involves borrowing in the usual sense of that word. These are (1) undistributed profits and depreciation allowances or so-called internal funds, (2) borrowing from banks or in the bond market, the source usually thought of in connection with the market rate of interest, and (3) equity financing or the sale of new stock issues. There are different costs and different risks incurred by the firm in using each of these sources, but the cost of funds from each may still be ex-

[11] E. Kuh and J. R. Meyer, "Investment, Liquidity and Monetary Policy," in *Impacts of Monetary Policy*, Commission on Money and Credit, Prentice-Hall, 1963, pp. 340–41. See also L. Tarshis, "The Elasticity of the Marginal Efficiency Function," in *American Economic Review*, Dec. 1961, pp. 958–85. For a survey of the econometric literature on the investment–interest rate relationship and a bibliography thereon, see M. J. Hamburger, "The Impact of Monetary Variables: A Survey of Recent Econometric Literature," in *Essays in Domestic and International Finance*, Federal Reserve Bank of New York, 1969, pp. 37–49, reprinted in W. L. Smith and R. L. Teigen, *Readings in Money, National Income, and Stabilization Policy*, 3rd ed., Irwin, 1974, pp. 376–90. For the results of a survey study on the effects of the rise in interest rates in 1966 and 1969–70, see J. Crockett, I. Friend, and H. Shavell, "The Impact of Monetary Stringency on Business Investment," in *Survey of Current Business*, Aug. 1967, pp. 10–26, and H. Shavell and J. T. Woodward, "The Impact of the 1969–70 Monetary Stringency on Business Investment," in the same publication, Dec. 1971, pp. 19–32.

pressed as a percentage so that one may be compared with the other.[12]

The funds that are lowest in cost to the firm are those derived from the first source, internal funds. The cost of using these funds for investment purposes is the opportunity cost measured by the rate that the firm could otherwise earn by putting these funds into securities. If the firm does not want to take the risk of purchasing stock, the appropriate rate is the rate it would earn on the purchase of debt issues of other firms or of the government. Because firms that are not in the lending business will usually have to pay a higher rate as borrowers than they can receive as lenders, the implicit cost to them of internal funds will be less than the cost of external funds secured by borrowing. Therefore, if we set out to build a supply curve showing the amount of funds available to the firm at different costs, we will show that the funds with the lowest cost are those internally generated. This amount, equal to retained profits and depreciation allowances for the time period, is shown in Figure 12-3 by segment *A* of the firm's marginal cost of funds schedule or MCF curve.

To secure a quantity of funds greater than that generated internally, the firm must resort to borrowing or to equity issues. Borrowed funds are shown in the figure by segment *B*. The upward slope of this segment indicates that the cost or the percentage rate paid for borrowed funds rises with the amount of borrowing. This is in part a reflection of the need to pay a higher rate the more heavily in debt the firm becomes within any time period, but more importantly it is the reflection of an imputed risk factor that rises with increased debt servicing.[13]

Since the cost of borrowing increases with the amount of borrowing, a point is reached at which it pays for the firm to resort to equity financing rather than to more debt financing. This is shown by the *C* segment, which lies above the *B* segment of the curve. The cost of equity funds is greater than the cost of borrowed funds for a number of reasons, one of the most important being the differential tax treatment of interest and dividends. The firm's interest payments are deductible in arriving at taxable income; its dividend payments are not. Because the federal corporate income tax rate is close to 50 percent, this in itself may make a dollar of equity funds almost twice as expensive as a dollar of borrowed funds. To illustrate, suppose a firm can sell additional bonds at a 6-percent rate. If it can sell additional stock only at a price that provides the purchaser with earnings per share equal to 6 percent of the price paid (i.e., at a 16.7 price–earnings multiple), it will have to earn about 12 percent on the capital goods financed by the stock issue to meet this 6-percent return, which is double the rate of earnings needed to meet the same 6-percent return on debt financing. A factor of a quite different nature that helps to explain the higher position of the equity-financing portion of the supply curve is an imputed cost in the form of the dilution of an existing group's control that results from an increase in the number of shares outstanding. Bond financing avoids this kind of cost. Apart from its higher average position, note also that the *C* segment of the curve slopes

[12]The approach to be sketched briefly in the following paragraphs is that developed by J. S. Duesenberry, in his *Business Cycles and Economic Growth*, McGraw-Hill, 1958, Ch. 5, and one that has since come to be widely used.

[13]The firm incurs debt to finance the acquisition of capital, which is expected to add to its profits after all expenses are met, including as an expense the interest on the debt incurred. However, the risk faced by the firm is that profits may not rise or may even decline, while interest costs will rise as debt rises. If profits do not rise, the firm can then meet the higher fixed cost of debt service in each time period only by cutting dividends, reducing the amount added to surplus, or incurring more debt to pay interest on the existing debt. All of these will have an adverse effect on the cost of raising funds in the future through both debt and equity, and thus involve risks that firms will take into account.

Applying this same argument, the *A* segment of the supply curve might also be drawn with an upward slope for firms with outstanding debt. Because debt is assumed to involve risk to the firm, the use of internal funds to finance investment entails rising costs inasmuch as these funds could alternatively be used to reduce the firm's outstanding debt.

upward as does the *B* portion. However, the upward slope of the *C* segment does not reflect imputed risk as in the case of borrowing, because the firm does not have to pay dividends but it does have to pay interest; what it reflects is a rising yield on the firm's stock (a declining price–earnings ratio), which will result from the fall in the market price of its stock as the firm issues more and more of it.

Various other considerations that underlie the shape of this curve could be noted, but it is sufficient for present purposes to see that a three-segment curve with the general shape of that in Figure 12-3 provides a reasonable description of how the cost of funds varies for the individual firm. The curve will differ considerably in shape and position from one firm to another but will in general have the appearance of that in the figure. When the separate curves for individual firms are aggregated, a smooth S-shaped curve like MCF_1 of Figure 12-4 will result.

Instead of the unrealistic assumption of a perfectly elastic or unlimited supply of funds at an ill-defined "the" interest rate, we now have an aggregate schedule that shows that an ever-larger total amount of funds is available only at successively higher cost. The concept of the interest rate still plays a major role here, for a rise or fall in the prevailing interest rate will shift the curve bodily upward or downward. A lower interest rate decreases the cost of both internal funds and equity funds, not merely the cost of borrowed funds. It is clear that the cost of internal funds is affected because this cost is measured by and varies directly with the market rate of interest. The cost of equity funds is affected because lower interest rates lead to a rise in the price of equities to restore the normal relationship between the interest rate on bonds and the price–earnings ratio on stock. A lower interest rate thus shifts the MCF curve as a whole downward—for example, from MCF_1 to MCF_2 in Figure 12-4.

The position of the MCF curve may also be altered with no change in the interest rate. An increase in corporate profits that results in an increase in retained earnings or a change in the method of charging depreciation that increases depreciation allowances during a time period would shift the supply curve to the right by elongating the elastic segment of the curve that

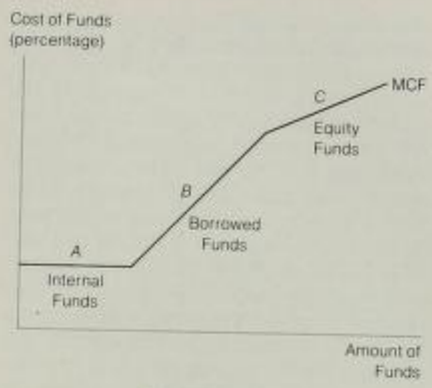


FIGURE 12-3
A firm's marginal cost of funds curve

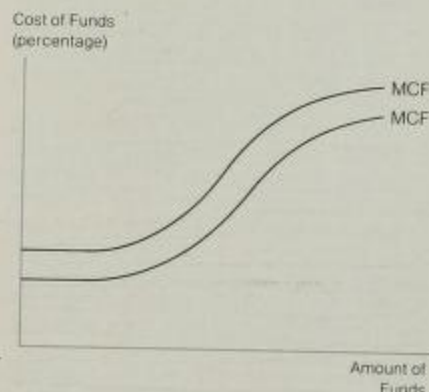


FIGURE 12-4
Aggregate marginal cost of funds curves

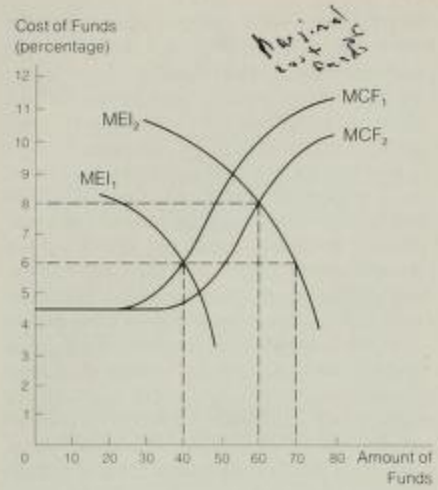


FIGURE 12-5
Shifts in the MEI and MCF schedules and the rate of investment

measures internal funds. The MCF curve would thus move from MCF_1 to MCF_2 in Figure 12-5. The leftward portion of the MCF curve remains unaffected—the new and old curves coincide over the elastic range of the old curve—but the rest of the curve shifts downward in much the same way that the curve shifts downward in response to a lower interest rate.

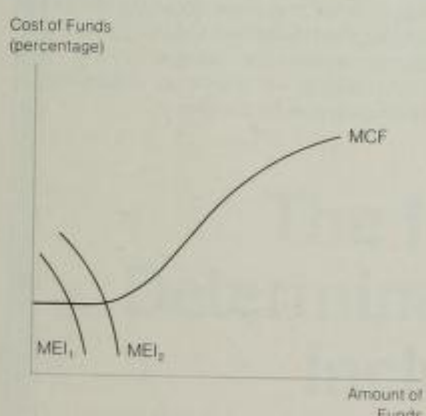
We can now see how shifts in either the MEI or the MCF curve or both affect the level of investment spending. Figure 12-5 illustrates one possible case. Suppose that the MEI curve shifts from MEI_1 to MEI_2, a movement that may be explained by the profits theory or the accelerator theory as in the preceding chapter. Whatever the cause of the rightward shift in the MEI curve, there is likely to be at the same time some increase in corporate profits, some increase in retained earnings, and thus some shift in the MCF curve, e.g., from MCF_1 to MCF_2. Assuming a greater shift in the MEI curve than in the MCF curve, which is most likely, the marginal cost of funds is driven up from 6 to 8 percent in this illustration. This rise in the cost of funds squeezes out some investment spending that otherwise would have occurred. If the cost of funds had remained at 6 percent instead of rising to 8 percent, investment would have risen to \$70 instead of being held to \$60.

The rise in investment spending would be even smaller if the interest rate increased at the same time. This would not be unreasonable in the present illustration; the expansion of economic activity that underlies the shift in the MEI curve would also work toward a higher interest rate. A rise in the interest rate, it will be recalled, will shift the whole MCF schedule upward, as from MCF_2 to MCF_1 in Figure 12-4. A shift of this kind in Figure 12-5 would move the MCF_2 curve upward to produce an intersection with MEI_2 at a cost of funds higher than 8 percent and a level of investment spending lower than \$60.

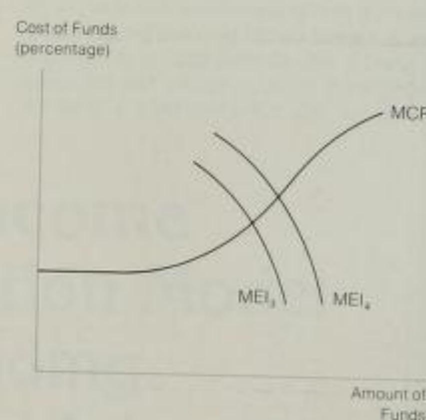
The results suggested by this illustration are plausible but not the only ones possible. As we have seen, the same economic forces that affect the position of the aggregate MEI curve also affect the position of the aggregate MCF curve, but the relative impacts may differ from time to time and give quite different results from time to time. In time of severe recession, the MEI curve may shift far to the left to a position like MEI_1 in Part A of Figure 12-6. At the same time, the elastic portion of the MCF curve will become shorter under recession conditions as retained profits shrink or possibly disappear. However, depreciation allowances will still provide an elastic range to this curve. Under these conditions, a shift in the MEI curve in the other direction that occurs as recovery gets under way need not run into an increase in the cost of funds. For example, the shift from MEI_1 to MEI_2 results in an increase in investment spending that is financed entirely out of internal funds. Thus, to the degree that this is the typical result during recession, it may be argued that the

interest rate is not an important financial determinant of investment during recession; it is retained earnings and depreciation allowances or internal funds that constitute the important financial determinant. On the other hand, as the economy moves toward a boom level, the MEI curve may intersect the MCF curve as shown by MEI_3 in Part B. Investment is now greater than can be financed out of internal funds; borrowing is taking place. At this stage, the interest rate becomes the important financial determinant of investment spending. The size of the particular increase in investment spending that results from the shift from MEI_3 to MEI_4 depends on the elasticity of the MCF curve in this range that covers borrowed funds. The interest rate is the price of borrowed funds and here becomes the important financial influence on investment.

The upshot of this is that we cannot evaluate the influence of finance on investment demand from information on the level of market interest rates alone. As we have just seen, changes in interest rates may be of little influence at some times and of considerable influence at other times. It would therefore clearly be more accurate to continue to take into account refinements on the side of finance such as the one here noted, but in later chapters we will find it necessary to return to the simplifications of earlier chapters. The whole complex of interest rates will once more be reduced to *the* interest rate, and the role of finance will once more be reduced to the cost of funds as measured by *the* interest rate. This becomes necessary if we are to use relatively simple models in our analysis in the following chapters. However, the reader should keep in mind the factors beyond the market interest rate that influence the level of investment spending from the side of finance and thus keep in mind the necessary qualifications to the conclusions suggested by the simple models that do not take their influence into account. The simple models will then be less misleading.



A Recession



B Prosperity

FIGURE 12-6
Comparison of possible effect of shift in the MEI schedule during recession and prosperity

A CONCLUDING NOTE

In Chapter 10 we worked our way through the pure mechanics of the process by which changes in the profit-maximizing or desired capital stock are translated into changes in the rate of investment spending. In Chapter 11 we saw how the profits and accelerator theories explain changes in the desired capital stock, and in the first part of this chapter we looked at the way a change in the interest rate does the same thing.

Whatever may have brought it about, a divergence in which the desired capital stock exceeds the actual capital stock will cause a rightward shift in the MEI curve. The greater the shift, the greater will be the rate of net investment spending. Once we have explained the shift in the MEI curve, we need look no further if we assume that the supply of funds available to finance investment is perfectly elastic at the prevailing market interest rate. However, as we saw in the last part of this chapter, the supply of funds is actually described by an upward-sloping MCF curve. It follows that any given shift in the MEI curve will have a larger or smaller effect on investment spending depending on the elasticity of the MCF curve in the relevant range. Although it is necessary in the following parts of this book to simplify by assuming that the role of finance is adequately indicated by a cost of funds curve that is perfectly elastic at the market interest rate, we saw in the foregoing pages that the role of finance does involve more than this.

In the following parts we will also have more to say about investment, but we will not enter further into the theory of investment behavior as such. In constructing models, we will, for the most part, assume changes in the rate of investment spending without considering in detail whether these changes resulted from the workings of the accelerator theory or any other theory. This corresponds to the procedure we followed in developing the first models in Part 2. With what we now know about the forces that actually affect investment spending, similar simplifying assumptions adopted in our later model-building will not be as misleading as would otherwise be the case.

part four

The Income Determination Model Including Money and Interest

chapter 13

The Aggregate Supply Function: Keynesian and Classical

The simple theory of income determination developed in Part 2 is called simple because the answers it provides to the core questions of what determines the level of income and output and the price level are based on highly simplified assumptions. Whether the economy is operating at the upper limit of output it is capable of producing with full employment of the labor force or at some position below that is made to depend entirely on aggregate demand. And because the simple theory assumes an aggregate supply curve that is perfectly elastic up to the full employment level of output, changes in the position of the aggregate demand curve over this range are without any effect on the price level. As was shown by Figure 4-1 on p. 48 and the whole set of figures that followed in Part 2, the level of output in each case depends solely on the position of the aggregate demand curve, and the price level is the same for every position of that curve up to the full employment level of output. In the simple theory of Part 2, the only role assigned to the aggregate supply curve is to set the price level; the higher the position of the curve the higher the price level, and vice versa. In that theory, output is demand-dominated and the price level is supply-dominated. Furthermore, in Part Two no attention was given to what determines the position of the aggregate supply curve or what makes it shift upward or downward. An initial position was arbitrarily selected and it was assumed that the curve did not shift from that position.

The major purpose of this chapter is to explain the derivation of the aggregate supply curve. As will be seen in this chapter, the curve itself takes three forms illustrated in Figure 13-1. Starting at the bottom, there is the curve in Part C: perfectly elastic up to the full employment level of output. This, as we saw in the chapters of Part 2, is the form of the curve on which the simple Keynesian model is based. We will see in this chapter that this curve, in turn, is based on the assumption that the money wage rate is fixed and the further assumption that the marginal product of labor is constant up to the full employment level of output. At the opposite extreme to the aggregate supply curve in Part C of Figure 13-1 is that in Part B: perfectly in-

elastic throughout and positioned at the full employment level of output. We will see in this chapter that such an aggregate supply curve follows from the assumption that the money wage rate is perfectly flexible downward, falling without limit in the event of an excess supply of labor. The model at the opposite extreme to the simple Keynesian model is the simple classical model, and it is the form of the aggregate supply curve shown in Part B on which the simple classical model is based. We will look at the simple classical model in some detail in the following chapter. Lastly, Part A is the intermediate case in which the aggregate supply curve becomes less than perfectly elastic at some level of output below the full employment level. We will see in this chapter that this curve follows from the assumptions that the money wage rate is fixed and that the marginal product of labor begins to diminish before the full employment level of output is reached.

Whatever its form, an aggregate supply curve is, in a sense, the summation of the supply curves of all the industries in the economy. To understand the nature of the aggregate supply curve, an approach which looks first at the individual firm's supply curve and the individual industry's supply curve proves helpful. In looking at these curves, attention will be limited to the conditions that yield a supply curve with an upward-sloping segment for the firm and the industry and that for all industries combined yield a supply curve like that in Part A of Figure 13-1. However, as already noted, we will give attention to the conditions that result in each of the three kinds of aggregate supply curves when we turn specifically to the derivation of the aggregate curve later in the chapter. We will discuss the cases shown in Parts A, B, and C of Figure 13-1 in that order.

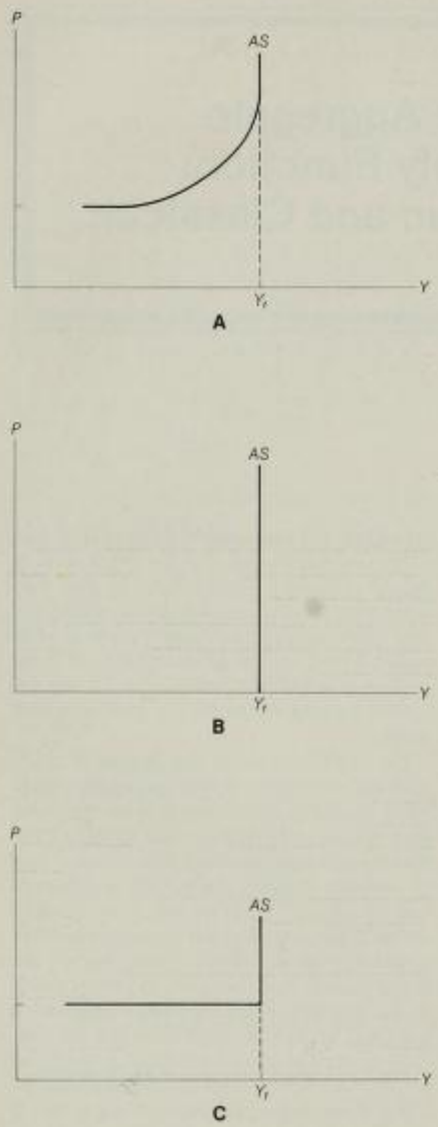


FIGURE 13-1
Aggregate supply functions

THE SUPPLY CURVE: FIRM AND INDUSTRY[1]

The supply curves we seek to derive are all short-run curves. Economists define the short-run period in the present context as a length of time too short to allow firms to vary the amount of plant and equipment with which they operate. If we assume that technology, or the method of production, also remains constant over the short-run period, the only way a firm can change its output in the short run is to change the amount of labor it employs. Thus, in short-run analysis, the relationship between a firm's labor input and goods output is the first consideration.[2]

This input–output relationship is described by the firm's production function, or total product curve, one possible form of which is shown in Part A of Figure 13-2. The linear portion may be said to indicate "proportional returns" because it indicates that output varies proportionally with labor input up to output of O_1 or up to labor input of L_1. However, because more and more labor, the variable input, is being employed with a fixed amount of plant and equipment as we move along the labor axis, the proportional relationship between output of product and input of labor must eventually be succeeded by one that is less than proportional. Thus we see that, beyond labor input of L_1, the increment to total output that results from each increment of labor input becomes smaller and smaller. The nonlinear portion of the total product curve, starting at L_1, illustrates this stage of diminishing returns to further labor input. Finally, beyond labor input of L_2, further labor input will not increase total output at all; in fact, at some even higher level of labor input, output will begin to decrease as more workers are hired. In Figure 13-2 we see that, at labor input of L_2, diminishing returns that begin at L_1 have diminished all the way to zero. Given the production function shown, it should be clear that in the short run the firm would under no circumstances hire labor beyond the amount of L_2.[3]

The increment in total product that results from the addition of a unit of labor is termed the *marginal physical product* (MPP) of that unit of labor. Part B of Figure 13-2 shows the curve of the marginal physical product of labor derived from the total product curve of Part A.[4] Since the total product curve shows that the increment in output is the same for each increment of labor input over the linear portion of that curve (i.e., up to labor input L_1), the MPP

[1]A minimal sketch of the relevant short-run micro theory is provided here. For a thorough treatment, see E. Mansfield, *Microeconomics: Theory and Application*, 2nd ed., Norton, 1975, Chs. 8–11, R. H. Leftwich, *The Price System and Resource Allocation*, 6th ed., Dryden Press, 1976, Chs. 8–13, or any other standard textbook on price theory.

[2]Except where we indicate otherwise, we will assume that all firms sell in purely competitive markets, that each firm attempts to maximize profits or minimize losses, that there is a constant money wage rate for each unit of labor provided by a perfectly homogeneous labor force whose services are sold in purely competitive markets, and that all firms are fully integrated. The last assumption means that, with the single exception of labor services, all variable inputs, including raw materials, necessary to the production of the firm's output are produced within the firm.

[3]Readers familiar with the orthodox theory of the firm will recognize the modification of that theory in the present construction. Conventional theory shows the total product curve of Part A divided into three stages: more than proportional, proportional, and less than proportional (or diminishing) returns. This means that the marginal product curve of Part B has an initial stage in which marginal product is rising, and the marginal cost curve of Part C has an initial stage in which marginal cost is falling. This first stage, however, does not appear in Figure 13-2. For present purposes, we may simplify without serious error by suppressing this first stage and showing only the two stages, proportional and less than proportional. It may be added that the firm, if it produces at all, produces an output greater than that in the first stage so that this stage is, in any event, of no practical interest.

[4]The vertical scale of Part B is an expanded version of that scale in Part A.

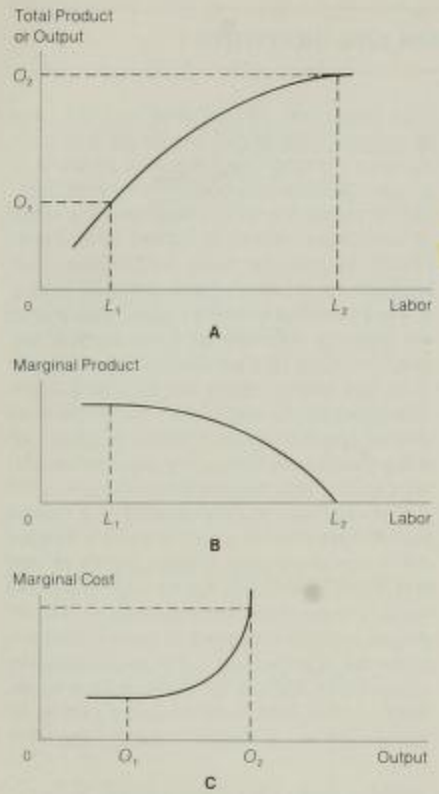


FIGURE 13-2
Derivation of the firm's supply curve

of labor is constant over this range of labor input. Between L_1 and L_2, however, the total product curve shows that the increments in total output become smaller and smaller and finally become zero. Therefore, the MPP of labor diminishes as labor input increases from L_1 to L_2 and finally becomes zero when labor input reaches L_2.

Since output may be increased or decreased in the short run only by increasing or decreasing the amount of labor employed, the only cost that varies with changes in output is the amount the firm pays for labor. The costs associated with a fixed plant and equipment are constant in the short run because they will be incurred by the firm whether the plant and equipment are being used to capacity or are sitting idle. Since *marginal cost* (MC) is the cost of producing an additional unit of output, the marginal cost of a unit of output in our model is the additional labor cost incurred in its production. Part C of Figure 13-2 gives the firm's MC curve, showing the increase in total cost resulting from each additional unit of output. Since MC in this case is composed entirely of labor cost, it equals at any level of output the money wage rate of labor, W, divided by labor's marginal physical product, MPP, at that level of output. Thus,

$$\text{MC} = \frac{W}{\text{MPP}}$$

Suppose, for example, that the total product curve shows that 10 worker-days of labor can turn out 100 units of total product and that 11 worker-days of labor can turn out 110 units. The MPP of the eleventh worker-day of labor would thus be 10 units. If the increase in total cost resulting from the labor cost of one worker-day of labor is \$20, the marginal cost of a unit of output over the range from 100 to 110 is accordingly \$2; that is, from the equation $\text{MC} = W/\text{MPP}$, $\$2 = \$20/10$. If the total product curve shows that 12 worker-days of labor can produce 120 units of output, the MPP of the twelfth worker-day of labor would also be 10, and the marginal cost of a unit of output over the range of output from 110 to 120 is again \$2 per unit. In terms of Figure 13-2, this indicates that labor input of 11 or 12 days must be below L_1, because only below L_1 does output vary proportionally with labor input. Since output varies proportionally with labor input up to L_1, the marginal cost of output must remain constant over the range of output that can be produced with labor input up to L_1. This range of output is zero to O_1, and the MC curve in

Part C accordingly shows the same marginal cost for all levels of output up to O_1.

If the firm were to expand output beyond O_1, the MPP of labor would start to decline, and the MC of output would start to rise. Suppose, for example, that when labor input is increased from 15 to 16 worker-days, total product rises from 145 to 150 units. The MPP of the sixteenth worker-day of labor is only 5 units, but the cost of the sixteenth worker-day of labor is, like all others, \$20. This means that the marginal cost of a unit of output over this range of 145 to 150 is \$20/5, or \$4. In terms of Figure 13-2, the sixteenth worker-day of labor must lie between L_1 and L_2, and the output from 145 to 150 must lie between O_1 and O_2. For this output, as we can see in Part C, marginal cost rises.

If, in its attempt to expand output, the firm should employ worker-days of labor of L_2 or more, the MPP of labor would be zero. Marginal cost would accordingly become infinite at output O_2. With the given MPP curve, labor clearly would not be employed beyond L_2 in the short run, no matter how low the wage rate of labor and no matter how high the market price of output, as long as profit remained the guide to the firm's employment decision.

In a purely competitive market, the individual firm is only one of so large a number of producers of a homogeneous product that no one firm has any control over the price at which this product can be sold. This market price is determined by industry supply and demand, and the individual firm simply adjusts its output to the market-determined price in order to maximize profits. The firm's short-run profit-maximizing level of output is the point at which the marginal cost of output, MC, just equals the market price of output, P.[5] Because of this and because at each possible level of output MC = W/MPP, the profit-maximizing output can also be expressed as that output at which

$$P = \frac{W}{\text{MPP}}$$

or that output at which

$$W = P \times \text{MPP}$$

This identifies the output at which the additional receipts from the sale of the additional output produced by an additional unit of labor (i.e., $P \times$ MPP) just equals the wage rate of labor.

For example, if the market price were \$2 per unit, output would be expanded to that level at which marginal cost was \$2. If the fixed wage rate were \$20 per worker-day, this would occur at that level of output at which the MPP of labor was 10. At this output, the MPP of labor, 10, multiplied by the price per unit of output, \$2, equals the wage rate of labor, \$20. If the market price of output were \$4, output would be carried to the higher level at which MC was \$4. With the same fixed wage rate of \$20 per worker-day, this higher level of output would be that at which the MPP of labor had fallen to 5. (At this output the MPP of labor, 5, times P, the price per unit of output of \$4, equals W, the wage rate.) Any level of output below that at which MC equals P—or, what is the same thing, at which receipts from the sale of the output produced by an additional unit of labor exceed the wage rate per unit of labor—would be a level of output whose expansion would increase profits. An expansion of output would add less to costs than to sales receipts. On the other hand, any level above the one at which MC equals P—or, what is the same thing, at which the receipts from the sale of the output produced by the last unit of labor are less than the wage rate per unit of labor—would be a level of output whose contraction would increase profits. A reduction in output would subtract more from costs than from sales receipts.

[5] Instead of MC = P, the more general expression is MC = MR, in which MR is *marginal revenue*, the additional revenue from the sale of one more unit of output. Under our assumption of purely competitive markets, each individual firm in the industry can sell the amount of output it wishes to at the going market price, so that MR for it always equals P. Under imperfect competition the firm can sell more only by cutting P, so that MR is always less than P. For simplicity only, our analysis is limited to purely competitive markets, in which the individual firm finds MR always equal to P.

The marginal cost curve of Part C of Figure 13-2, which is derived entirely from the MPP of labor and the wage rate of labor, is also the firm's supply curve. It shows the various quantities of output (measured along the horizontal axis) that the firm will supply to the market at each price (measured along the vertical axis) in order to maximize its profits.[6]

Once the profit-maximizing level of output has been determined, the amount of labor that should be hired by the firm in order to achieve this level may be found directly from the total product curve given in Part A of Figure 13-2. Given our other assumptions—as long as the money wage rate, the curve of the marginal product of labor, and the market price of output remain unchanged—the short-run equilibrium level of output and the amount of labor employed by the individual firm will remain unchanged. However, anything that reduces the money wage rate, alters the total-product curve in a way that raises the marginal product of labor, or raises the market price at which output can be sold will increase the level of output at which the firm maximizes profits. The amount of labor to be hired at this new level of output may, as before, be determined graphically from the appropriate point on the firm's total product curve.

The relationships we have just described for the individual firm apply to all firms in a purely competitive industry. The industry's supply curve is then simply the horizontal summation of the marginal cost curves of all firms in the industry, and, as shown by Figure 13-3, it looks much like the firm's supply curve.[7] In Figure 13-3 the vertical axis is now labeled P to make clear that a supply curve shows the amount of the commodity supplied at each possible price for the commodity.

Since each firm in the industry operates in the short run with a fixed stock of plant and equipment and a fixed technology, the industry also operates with these fixed factors in the short run. For the same reason that the firm's supply curve eventually slopes upward to the right, the industry's supply curve must eventually slope upward to the right. Given these assumptions, beyond some point more output can be produced by the industry only at higher marginal cost and will therefore be supplied only at higher prices.

THE SUPPLY CURVE: AGGREGATE

To proceed from the derivation of the supply curve for a single firm and single industry to that for industry as a whole is essentially a matter of extending and, in some cases, modifying the theory sketched for the firm and the industry. The following pages will present the derivation of the three forms of the aggregate supply curve shown in Figure 13-1. The first section derives the curve in Part A. This curve, which slopes upward to the right, is based on the assumptions of a fixed money wage rate and diminishing marginal productivity of labor. As these are

[6]Since MC is constant up to O_1, the firm's average variable cost is constant over this same range of output. If price is above P_1, the firm maximizing profits (or minimizing losses) will produce and sell something more than O_1. However, if price is below P_1, the firm will produce and sell no output, for this is a price that does not even cover its variable cost (here, labor cost) per unit of output. If price were exactly P_1, the firm would incur a loss equal to its fixed costs, whether it produced no output or output O_1.

[7]Although this curve is similar in shape to the individual firm's supply curve in Part C of Figure 13-2, it must be recognized that the scale of the two figures differs on the horizontal but not on the vertical axis. A given interval along the horizontal axis of the industry graph represents a much larger number of units of output than the same interval along the axis of the individual firm's graph.

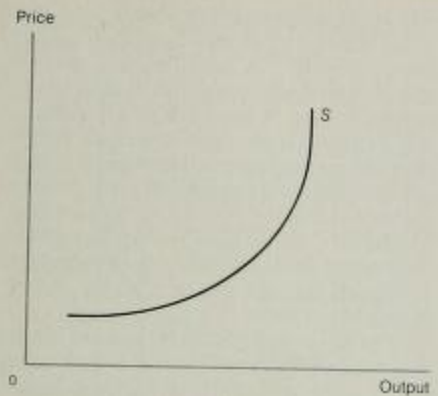


FIGURE 13-3
Industry supply curve

the same standard assumptions that underlie the short-run supply curve for the firm and industry, the relationship between output as a whole and the general price level shown in Part A of Figure 13-1 is not unexpectedly similar to that shown in Figure 13-3 for the single industry. The second section derives the perfectly inelastic aggregate supply curve or that in Part B of Figure 13-1. This curve is based on the assumptions of a perfectly flexible money wage rate and, like that in Part A, a diminishing marginal productivity of labor. Finally, the third section derives the aggregate supply curve, which is perfectly elastic up to the full employment level of output or that in Part C of Figure 13-1. This curve is based on the assumption of a fixed money wage rate as is the curve in Part A, but, unlike the curve in Part A, it is based on the assumption of constant marginal productivity of labor.

In each of the three cases, the derivation will be presented graphically through a four-part diagram adjusted as required to fit different cases.

A. The Upward-Sloping Aggregate Supply Curve

We may begin with Part A of the four-part diagram in Figure 13-4. Part A is the aggregate production function, or total product curve, for the economy as a whole. Apart from the fact that the axes have been reversed for convenience, the curve has the same appearance and the same properties as the individual firm's production function, or total product curve, given in Part A of Figure 13-2. If the composition of aggregate output is fairly stable, the aggregate production function will show a range of proportional returns followed by a range of diminishing returns for essentially the same reasons that the production function of the individual firm does. The attempt of a single firm or a single industry to expand its output further and further runs into the short-run barrier of fixed plant and equipment; the attempt of all firms and industries to expand aggregate output further and further runs into the same barrier.

Part B gives us the curve of the marginal physical product of labor which is now labeled D_N for reasons that will be noted. As in Part B of Figure 13-2, this curve is derived directly from the production function in Part A. For the uses to be made of Figure 13-4, the axes of Part B have also been reversed from those in Figure 13-2, but the meaning of Part B is the same in both figures. The MPP schedule for the individual firm indicates the additional output that results from the addition of each unit of labor input. Similarly, the MPP schedule for industry as a whole indicates the addition to the economy's aggregate output that results from the addition of each unit of labor input.[8]

Due to diminishing returns, the MPP of labor decreases as we move up the vertical axis of

[8]In practice, of course, we would be dealing in units of thousands of workers when estimating actual changes in the labor input for the economy as a whole, but the same principle applies in both cases.

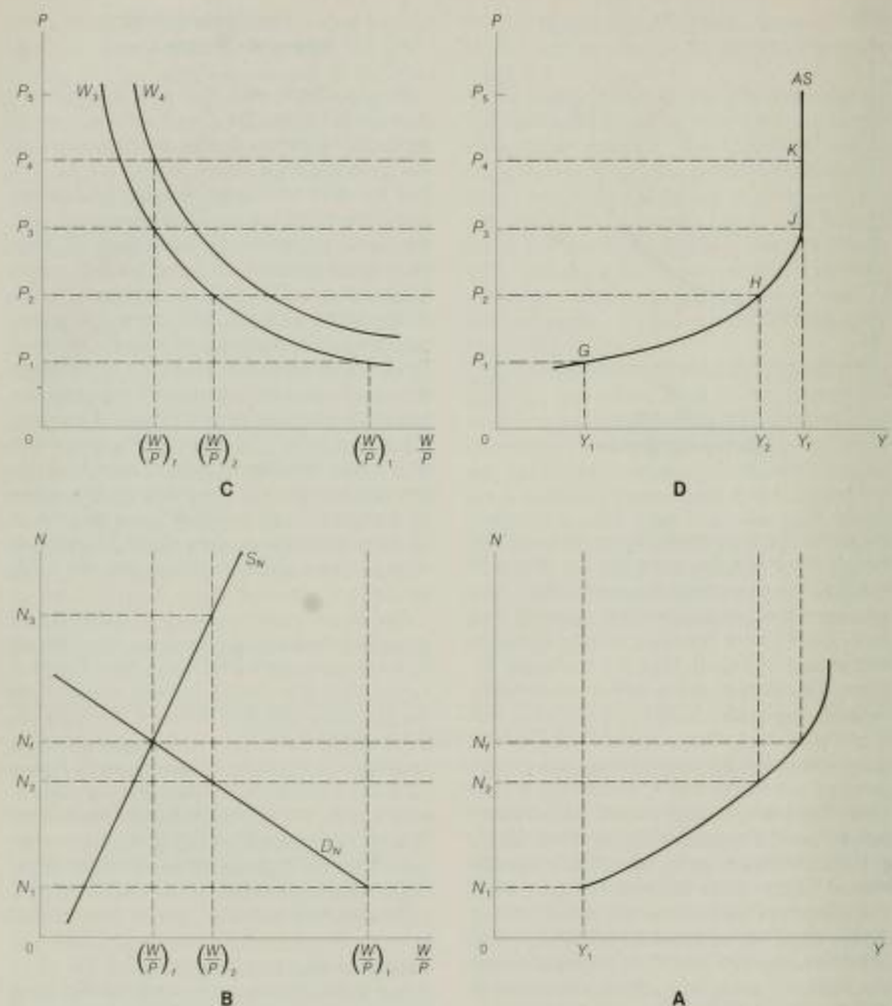


FIGURE 13-4
Derivation of the aggregate supply curve: fixed money wage rate

Part B to larger amounts of labor. Since employers maximize profits by hiring labor up to the point at which MPP $\times P = W$, for any given W more labor will be hired beyond an initial equilibrium only at a higher P, or for any given P more labor will be hired only at a lower W. The level of employment consistent with profit maximization may alternatively be described as the level at which MPP $= W/P$ or the level at which MPP equals the real wage (which is the ratio of the money wage to the price level). For an illustration, suppose initially that the amount of labor employers hire is N_1 which means that W and P are such that W/P equals the MPP of labor at that level of employment. With the values of both W and P assumed given, to hire more or less than this number of workers would sacrifice maximum profits, and employers would not do so. Under what conditions would they be willing to expand the number of workers employed to, say, N_2? To do this is to encounter the lower MPP of labor at N_2, so employers will only expand N to N_2 if W/P declines by the required degree. Thus, we may say that the amount of labor hired is an inverse function of the real wage. The lower the real wage, the larger the number of workers employed. Viewed in this way, the MPP curve becomes a demand curve for labor and for this reason is labeled D_N in Part B.

The other curve in Part B, which is labeled S_N, is the supply curve of labor. The model shows the supply of labor, like the demand for labor, to be a function of the real wage, W/P, with the supply of labor a direct function. This expresses the contention that a higher money wage rate will not call forth more labor if the price level rises proportionally, for in this event the real wage remains unchanged. For the same reason, a lower money wage rate matched by a proportional fall in the price level will not lead to a reduction in the quantity of labor supplied. The basis for this postulate of classical theory is the unpleasantness or irksomeness of more work; a larger *real* reward is necessary to induce labor to provide an ever larger supply of labor services.

Workers, as well as the firms that employ them, are maximizing units in this system—workers seek to maximize utility just as firms seek to maximize profits. Firms will not hire more labor at a lower money wage rate if the prices at which they can sell their output fall proportionally with the money wage rate. Of relevance to the firm is the cost of a unit of labor relative to the price at which the firm's output sells—it is the real wage that counts. By the same token, workers will not supply more labor at a higher money wage rate if the prices of the goods purchased with their wages rise proportionally with the money wage rate. Of relevance to the worker is the money wage received per unit of labor supplied relative to the prices of the goods that can be purchased with that money wage—it is the real wage that counts. This maximizing behavior on either side of the market for labor gives us the demand curve for labor as an inverse function of the real wage and the supply curve as a direct function of the real wage.

Turning now to Part C, this does nothing more than depict graphically a relationship covered in earlier discussions of the theory of the firm and industry. Recall that, in its attempt to maximize profits, the single firm will in the short run expand output to the level at which MC $= P$. Since MC $= W/$MPP, the profit-maximizing output is also that at which $P = W/$MPP. Thus, for fixed money wage rate, W, the firm will hire labor up to the amount at which labor's MPP multiplied by P equals the given W. For the single firm and, by the same argument, for the single industry, once the profit-maximizing output has been determined for the fixed wage rate, more labor will be hired and more output produced only if forces within the market should produce a rise in P. The amount of labor employed by the firm could then expand until the MPP of labor (which will decrease as the amount of labor employed increases) multiplied

by the new, higher P is equal to the unchanged W.

The same follows for industry as a whole. Curve W_3 in Part C of Figure 13-4 gives us nothing more than the various combinations of P and labor's MPP whose products equal the W that was assumed in drawing the fixed W_3 curve. For example, if the level of employment N were such that the MPP of labor was 8, P would have to be 3 before producers would provide this level of employment with W fixed at 24. If N were greater, so that MPP was smaller—for example 6—P would have to be 4 in order to make the employment of labor up to the amount at which its MPP fell to 6 consistent with the profit-maximization objective of producers.[9]

Part D rounds out the apparatus by showing how the aggregate supply function is derived from the other parts of the figure. Thus, for any level of N, Part A gives the aggregate output, Y, corresponding to that labor input, but, as we will see below, only the levels of N up to N_f are relevant. The D_N curve in Part B indicates the real wage, W/P, equal to the MPP of labor, for each possible level of N. And the W_3 curve of Part C identifies the price level, P, which is required to yield the indicated W/P on the basis of the money wage rate of W_3. The P indicated in Part C and the Y indicated in Part A then combine to determine a specific point in Part D that is one point on the aggregate supply function being derived. Such a specific point may also be determined by starting in Part C. Assuming W is given at W_3, for a selected level of P the W_3 curve identifies a resultant W/P on the horizontal axis. The D_N curve in Part B then shows the amount of labor that will be employed at that W/P. And the production function in Part A shows the amount of output or Y produced by that amount of labor. The P assumed in Part C and the Y found in Part A then combine to identify a specific point in Part D.

As an illustration, start in Part A with N of N_1, an amount of labor that produces output of Y_1. The D_N curve in Part B shows the MPP of labor and therefore the W/P at which that quantity of labor will be demanded, namely, $(W/P)_1$. On the assumption of a fixed money wage rate, specifically W_3 in Part C, producers must get a price level of P_1 for their output of Y_1 to yield $(W/P)_1$, the W/P at which they will hire N_1 and produce Y_1. Therefore, producers will supply the indicated amount of output, i.e., Y_1, at the price level of P_1. The P_1 in Part C and the Y_1 in Part A combine to identify the specific point G in Part D which is one point on the aggregate supply curve.

In the same way, other values of N up to N_f, e.g., N_2 and N_f, will identify other points, H and J, respectively, in Part D which will be other points on the aggregate supply curve. Connecting such points as G, H, and J yields an aggregate supply curve that slopes upward to the right. Up to the limit of N_f, the greater N, the greater will be the Y and P indicated in Part D. A greater N will mean a greater Y as more labor produces more output up to the N at which the MPP of labor falls to zero, but a greater N must also be accompanied by a higher P. Given the assumption that W is fixed, a higher P is the only way to obtain a lower W/P, and, because the higher N means a lower MPP, employers will not find it profitable to hire the additional labor except at the correspondingly lower W/P.

What must next be seen is that once output reaches the level Y_f which is the level produced with employment at N_f, the aggregate supply curve no longer slopes upward to the right but becomes vertical. Output has reached the maximum set by full employment of the labor force,

[9] The W_3 curve is one of a family of such curves that will be employed in Part C of a series of figures like Figure 13-4. Technically, each of the curves in question is a rectangular or equilateral hyperbola which is a curve such that all rectangles established by running perpendiculars from any point on the graph of the function to the axes will be equal in area. The general equation for this function is $xy = a$ in which $a > 0$. For the variables here, the specific equation becomes $(W/P)P = W$, and in the two illustrations given, $8 \times 3 = 24$ and $6 \times 4 = 24$.

the level of employment identified by the intersection of the S_N and D_N curves in Part B.[10] At the market-determined real wage rate established by this intersection, every worker who wishes to work at that real wage is able to find employment; the level of employment at this real wage is, therefore, full employment. However, take any level of employment below this such as N_2 in the illustration above. As shown by the D_N curve, N_2 will be employed at a real wage of $(W/P)_2$, but the S_N curve shows that N_3 will be available at that real wage rate. There is an excess supply of labor in the amount $N_3 - N_2$. Similarly, there is an excess supply at each lower real wage rate down to $(W/P)_f$ at which rate the supply of labor equals the demand for labor. Because the money wage rate is assumed to be fixed in the present case, the only way to move from a higher real wage rate to the lower rate which will raise employment to N_f and output to Y_f is through a rise in the price level to P_3. As P rises toward P_3 with W stable at W_3, W/P falls and employment rises toward N_f.[11]

Rises in P beyond P_3 with a fixed W will mean declines in W/P below $(W/P)_f$. It might at first appear that the amount of labor employed will then fall below N_f because the S_N curve shows that less than N_f will be provided at a real wage below $(W/P)_f$. This would be true if the money wage rate were absolutely fixed. However, the notion of a fixed money wage rate is understood to preclude only downward adjustments, not upward adjustments. Or it is understood to be inflexible downward but not inflexible upward. The idea of a fixed money wage rate reflects labor's opposition to a cut in that rate but labor obviously has no opposition to a rise in its money wage rate.

What then happens if P rises above P_3 is that W rises proportionally with P. Such a rise in P with no rise in W would create an excess demand for labor as the real wage rate fell below $(W/P)_f$. The quantity of labor employers sought to hire would exceed the amount provided and competition would bid up W to match each rise in P. For example, a rise in P from P_3 to P_4 would raise W proportionally as shown by the shift in the W curve from W_3 to W_4 in Part C. P at P_4 and the W curve at W_4 would keep W/P at $(W/P)_f$, N at N_f, and Y at Y_f. Thus, in Part D of Figure 13-4, for any value of P above P_3, Y remains at Y_f. The points J and K show the same Y, i.e., Y_f, or the aggregate supply curve becomes perfectly inelastic upward starting at J.

B. The Perfectly Inelastic Aggregate Supply Curve

In a world with no labor unions or other labor market imperfections, the money wage rate would be flexible downward. Any time that the supply of labor exceeded the demand for labor, competition among workers for jobs would result over time in a decline in the money wage rate. In this case, the aggregate supply curve would turn out to be perfectly inelastic at the full employment level of output like the curve shown in Part B of Figure 13-1.

The derivation of the aggregate supply curve on the assumption of a flexible money wage rate is traced in Figure 13-5. Apart from the two

[10] Note that this is not the maximum output the economy is capable of producing. Part A shows that output would continue to expand beyond Y_f with labor input greater than N_f. The existing stock of capital and state of technology are such that labor input greater than N_f still has a positive marginal physical productivity. Thus, the aggregate supply curve in Part C becomes vertical before the production function in Part A does.

[11] There is no force that automatically raises P toward P_3, and therefore N toward N_f. It should be apparent that what P will be cannot be determined until we combine an aggregate demand curve with the aggregate supply curve in Part D. As will be covered in detail in Chapter 17, there can be an equilibrium in the present model, for example, with P of P_2, output of Y_2, employment of N_2, and unemployment of $N_3 - N_2$, if the aggregate demand curve happens to intersect the aggregate supply curve at the point H. The economy would be in equilibrium with less than full employment. The aggregate supply curve identifies the P needed to provide the Y consistent with full employment as well as the P consistent with other levels of Y in the present case, but it does not and can not by itself determine what the actual level of P, Y, or N will be.

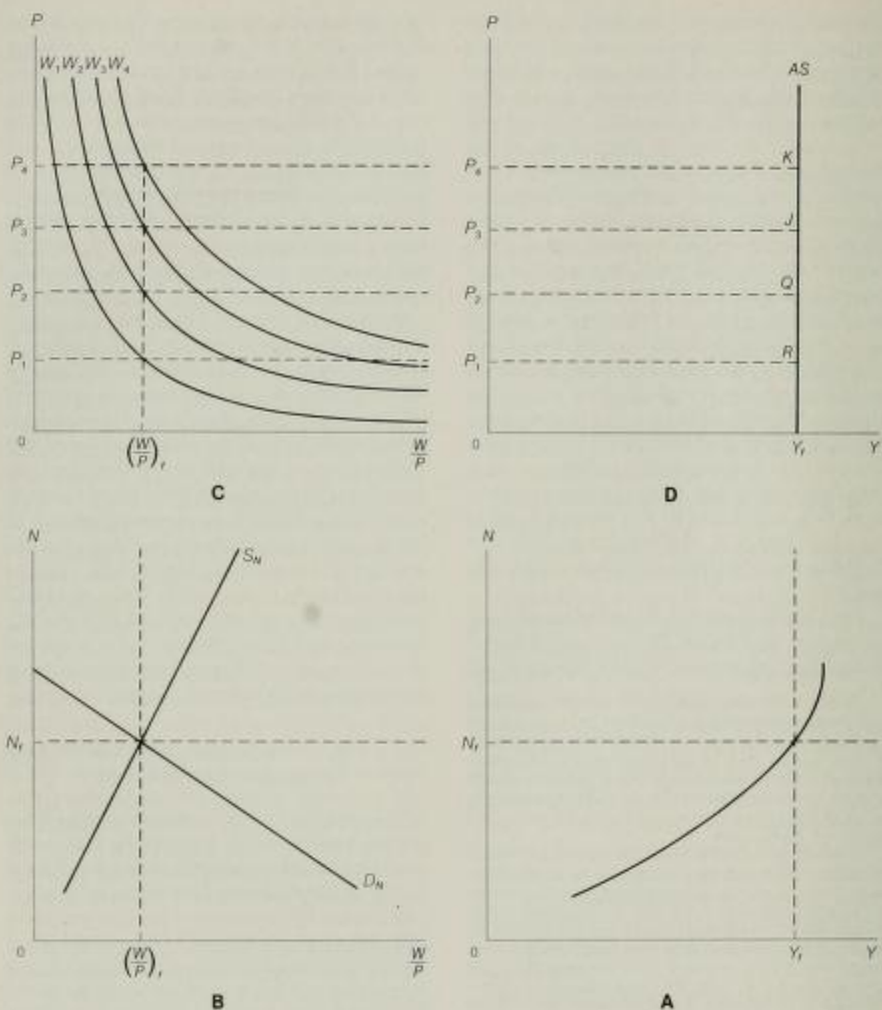


FIGURE 13-5
Derivation of the aggregate supply curve: flexible money wage rate

additional curves in Part C to designate two more possible money wage rates, W_1 and W_2. Parts A, B, and C of Figure 13-5 are identical with those parts of Figure 13-4. The substantive difference is in Part D; the aggregate supply curve is perfectly inelastic throughout. To see the basis for this difference, start off with the assumption that the money wage rate is W_3 in Part C. On this assumption, to obtain the W/P consistent with full employment, N_f, calls for P of P_3 as P_3 and W_3 give $(W/P)_f$. Y_f in Part A and P_3 in Part C combine to identify the point J in Part D. This is no different from the derivation of point J in the preceding part of this chapter.

Assume next that P falls to P_2. On the assumption of a fixed money wage rate, we found in Figure 13-4 that N would fall to N_2. The decline in employment would mean a decline in output or at P of P_2 firms would produce output of Y_2 as shown in Figure 13-4 by the point H. On the assumption of a flexible money wage rate, the result is quite different. W will not remain at W_3 as P falls to P_2 because this will create an excess supply of labor, and a flexible money wage rate is one that falls in the face of an excess supply of labor as workers compete for the available jobs. Thus, the money wage rate will decline from W_3 to W_2 in Part C of Figure 13-5. This decline in W is proportional to the decline in P so the preexisting real wage of $(W/P)_f$ is maintained. At that real wage, N is N_f and Y is Y_f. Combining Y of Y_f in Part A with P of P_2 in Part C yields the point Q in Part D. The only change is that P and W have fallen proportionally. In the same way, a further decline in P to P_1 would bring about a decline in W to W_1. Therefore, W/P remains at $(W/P)_f$, N at N_f, Y at Y_f, and the combination of Y_f and P_1 yields the point R in Part D. Again, the only change is that W and P have fallen proportionally. With a downwardly flexible money wage rate, the aggregate supply curve turns out to be perfectly inelastic at the full employment level of output. Whatever the level of P, W adjusts to it to maintain W/P at $(W/P)_f$, thereby maintaining employment at N_f, and output, in turn, at Y_f. The assumption of a flexible money wage rate fixes the level of total production at Y_f for it makes the level of total production independent of the price level.

C. The Perfectly Elastic Aggregate Supply Curve

The perfectly elastic aggregate supply curve in Part C of Figure 13-1, like the curve in Part A which slopes upward to the right, assumes a fixed money wage rate. However, what makes it perfectly elastic instead of merely elastic like the curve in A is the additional assumption of a constant marginal productivity of labor. Part A of Figure 13-6 shows a production function that gives this result. Over a range of output up to and beyond that which can be produced with full employment of the labor force, the production function is perfectly linear. Each increment of labor employed produces the same increment to output so that the curve of the marginal product of labor in Part B is at the same height above the employment axis over the relevant range of employment. For reasons earlier explained, the marginal product curve of labor is the demand curve for the labor, D_N, and this curve intersects the S_N curve to identify the full employment position, N_f. The S_N curve here is the same as in the preceding cases.

Assume an initial full employment equilibrium which occurs with W/P of $(W/P)_f$. N is N_f and Y in Part A is then Y_f. Assuming the fixed money wage rate given by W_3 in Part C, to secure W/P of $(W/P)_f$ requires P of P_3 in Part C. The combination of P_3 and Y_f identifies the point J in Part D. Since this initial equilibrium is at full employment, a rise in P from P_3 to P_4 will give us the same result found in the preceding cases. As has been noted, the notion of a fixed money wage rate rules out decreases but not increases, and the effect of a rise in P is a rise in W proportional to the rise in P. W shifts to W_4; W/P, N, and Y all remain unchanged. P_4 and Y_f indicate

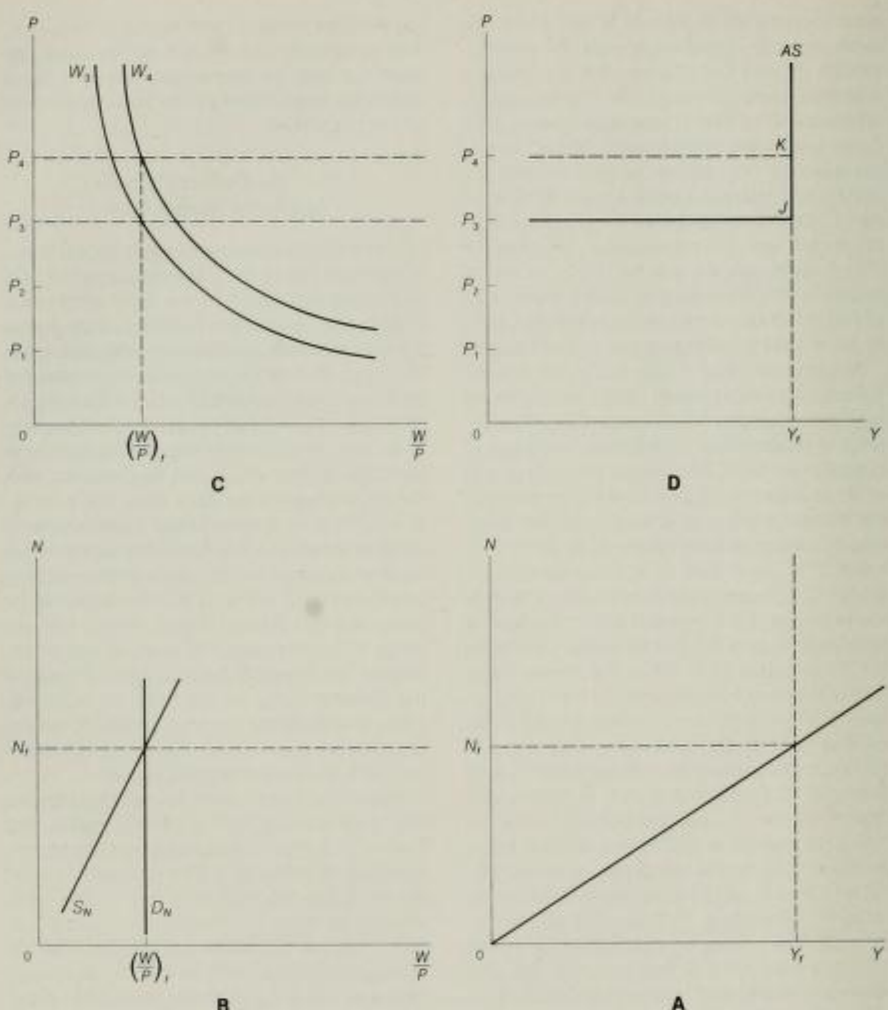


FIGURE 13-6
Derivation of the aggregate supply curve:
fixed money wage rate and
constant marginal productivity of labor

the point K on the aggregate supply curve. As before, the aggregate supply curve becomes perfectly inelastic at the full employment level of output.

Again, starting at the full employment equilibrium, assume now a decline in P from P_3 to P_2. Because the D_N curve is perfectly elastic at $(W/P)_1$, the slightest decrease in P with W fixed makes the total amount of labor supplied at the resultant higher W/P entirely an excess supply. Employers will hire any amount of labor over the range shown in Part B at the real wage rate of $(W/P)_1$, but they will hire none at all at any real wage above this. Hence, with W perfectly rigid in the downward direction, the slightest decline in P below P_3 will reduce N to zero and therefore reduce Y to zero. Graphically, this is shown by a perfectly elastic aggregate supply curve positioned at the P_3 level.

An aggregate supply curve like this is plainly a limiting case. The aggregate production function in practice may be approximately linear and therefore the marginal product of labor may be approximately constant at relatively low levels of employment and output. As additional workers are hired over this limited range, the existing stock of capital is so large relative to the still low level of employment that each additional worker can be supplied with as much as he can effectively work with and therefore his marginal product will not fall below that of workers hired earlier. In this event, the aggregate supply curve may be expected to have a perfectly elastic segment at low levels of output. However, as the economy moves toward full employment, the amount of capital available per worker is no longer the amount that enables each additional worker to produce the same additional output as was the case with fewer workers employed. The marginal product of labor will begin to decline or diminishing returns will set in. This will result in the segment of the aggregate supply curve which slopes upward to the right or the segment which was derived as the first of the three cases considered here.

On the basis of this observation, the simple Keynesian theory of income determination developed in Part Two is seen to fit best an economy producing output far below the full employment level. It has long been recognized that the simple Keynesian model developed in Part Two applies strictly during times of depression, one characteristic of which is a level of output far below the full employment level. The aggregate supply curve does have a range way down there in which it is perfectly elastic or very nearly so, and if we happen to be operating within that range, the simple Keynesian model is not as much of an oversimplification as is otherwise the case. However, where a great oversimplification does enter is with the assumption that the aggregate supply curve has this perfectly elastic shape all the way to the full employment position. That was the assumption adopted in Part Two to get the analysis off the ground in the easiest way; our later work will take into account that, apart from a range at very low levels of output, the aggregate supply function will slope upward to the right under conditions of a fixed money wage rate.

A CONCLUDING NOTE

The derivation of the aggregate supply curve in this chapter has shown how the aggregate amount of output supplied by industry as a whole varies with the price level in each of the three cases examined. The amount supplied at each price level, will, of course, remain as shown by any given aggregate supply curve as long as the determinants of the curve remain unchanged. However, as the determinants do change over time, the aggregate supply curve

will shift accordingly. Thus, a rightward or outward shift in the production function in Part A of Figures 13-4 to 13-6, something which can and does occur in the long run, will produce an outward shift or an increase in aggregate supply in Part D of those figures. A leftward shift or increase in the labor supply in Part B will have the same effect on the vertical portion of the aggregate supply curve. Also, in Figures 13-4 and 13-6, which are based on the assumption of a fixed money wage rate, there will be an upward shift or a decrease in the non-vertical portion of the aggregate supply curve as a result of an autonomous increase in the money wage rate that fixes it at a new higher level. A shift cannot occur for this reason in Figure 13-5, as that is based on the assumption of a perfectly flexible money wage rate.

The mechanics of the process by which a shift of the curves in Parts A, B, and C brings about shifts in the aggregate supply curve could have been traced as a final part of this chapter. However, it is not very revealing to trace the process by which such shifts occur in isolation from aggregate demand. If one traces this process in a framework that contains an aggregate demand curve, he thereby not only sees how the forces in question exert their influence on aggregate supply but also on the output level and the price level. In the following chapter, which examines the simple classical theory of income determination, we will see how shifts in the production function and in the supply of labor displace the perfectly inelastic aggregate supply curve that is a part of that theory and, in turn, how such displacement of the aggregate supply curve in combination with an aggregate demand curve based on classical theory brings about changes in the output level and the price level. In later chapters we will extend the Keynesian theory of aggregate demand developed in simple form in Part Two and see how shifts in the aggregate supply curve in combination with an aggregate demand curve based on Keynesian theory brings about changes in the output level and the price level.

chapter

14 The Basic Classical Model

Keynes: Aggregate demand determines output
Agg supply determines price

Classical

In the simple Keynesian theory of income determination covered in Part 2, we saw that the level of output is determined solely by aggregate demand. The perfectly elastic aggregate supply curve which is part of that theory is without influence on the level of output; it determines only the price level. Having introduced the concepts of the production function, the demand for and supply of labor, and wage flexibility in the preceding chapter, we now have the basis on which to develop the simple classical theory of income determination, a theory that stands at the opposite extreme to the simple Keynesian theory. In the classical theory, the level of output is determined solely by aggregate supply, just the reverse of the result found in the simple Keynesian theory. Furthermore, in the simple classical theory, aggregate demand determines only the price level, again just the reverse of the result found in the simple Keynesian theory.

What was done in Part 2 for the simple Keynesian theory will be done here for the simple classical theory, albeit in much shorter form. Although the model to be developed is an abbreviated version of the classical one, the form in which it is presented will nowhere be found in the writings of Ricardo, Mill, Marshall, or any of the other nineteenth and early twentieth century economists who were the creators and refiners of the classical theory. It was not until after the appearance of Keynes's *General Theory*[1] that economists set to work to construct from the writings of the classical theorists complete models that might be placed side by side with the Keynesian model for comparison. Since, for simplicity, we will limit ourselves to the broad outlines of that classical model, we must note that our inattention to specific details amounts to inattention to the many qualifications that must be made to the conclusions we will draw. Nonetheless, the conclusions reached on this basis will at least suggest the directions in which the classical theory led.

The classical theory, whose broad outlines are described in this chapter, is specifically classical theory as it was before Keynes's *General Theory*. As we will see below, the quantity theory of money is an essential part of classical theory, but today there is a "modern" version of the quantity theory that differs markedly from the old version that was part of traditional clas-

[1] John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936.

sical theory. We will touch on the modern version in a later chapter; here our purpose is to cover the essentials of classical theory before Keynes, and so it is the "old" quantity theory that is examined as a part of that theory.

Our classical model is also limited to the classical explanations of the determination of employment, output, and price levels, for it was these explanations that came under attack by Keynes and others. The other basic questions of what is to be produced, how it is to be produced, and for whom it is to be produced provoked no disagreement. In Keynes's words:

> If we suppose the volume of output to be given, i.e. to be determined by forces outside the classical scheme of thought, then there is no objection to be raised against the classical analysis of the manner in which private self-interest will determine what in particular is produced, in what proportions the factors of production will be combined to produce it, and how the value of the final product will be distributed between them.[2]

marginal physical product

OUTPUT AND EMPLOYMENT IN CLASSICAL THEORY

The equilibrium levels of output and employment are determined in the classical system as soon as we are given (1) the economy's production function, from which is derived the demand curve for labor, and (2) the supply curve of labor. Apart from the inclusion in Part D of aggregate demand curves, AD_0 and AD_1, which will be explained below, all of the other curves in Figure 14-1 were explained in deriving Figure 13-5 in the preceding chapter. The MPP of labor curve given by D_N in Part B is as before derived from the production function in Part A and Figure 14-1 shows the standard case of diminishing returns in which the MPP of labor decreases as the amount of labor employed increases. As was earlier explained, employers hire that amount of labor at which MPP $= W/P$, and therefore the amount of labor hired is an inverse function of the real wage. With a given production function and an initial equilibrium at which MPP $= W/P$, an increase in employment is not possible without a decrease in the real wage. As before, the curve labeled S_N in Part B is the supply curve of labor. This shows that the amount of labor supplied is a direct function of the real wage rate, the reasons for which were also considered earlier.

On the assumptions underlying the classical model, the intersection of the supply and demand curves for labor in Part B determines what the level of aggregate output will be. With a real wage of $(W/P)_0$ in Figure 14-1, there is equilibrium between the supply of and demand for labor. At this real wage, employers choose to hire N_0 of labor, and workers choose to provide N_0 of labor. With the aggregate production function of Part A, employment of N_0 indicates aggregate output of Y_0.[3]

In the classical scheme of things, any real wage other than $(W/P)_0$ will generate forces causing the real wage to rise or fall by the

[2]Keynes, *General Theory*, pp. 378–79. This is not to say, however, that the theory of distribution that was developed by classical economists is the last word. Although Keynes accepted this part of classical theory for his purposes, the very ground broken by Keynes in the theory of aggregate output and employment gave rise to a new interest in the macroeconomic aspects of income distribution. See, for example, S. Weintraub, *A General Theory of the Price Level, Output, Income, Distribution, and Economic Growth*, Chilton, 1959, and N. Kaldor, "Alternative Theories of Distribution," in *Review of Economic Studies*, Vol. 23, 1955–56.

[3]Although N_0 and Y_0 are full employment values, they have not been labeled N_f and Y_f as in the preceding chapter. Shifts in the production function and in the supply of labor are introduced below, and there is a different full employment value for N and Y for each position of these curves. For this reason, each value for N and Y, whether a full employment value or not, has been simply identified by a numerical subscript. The original set of equilibrium values is identified by the subscript 0.

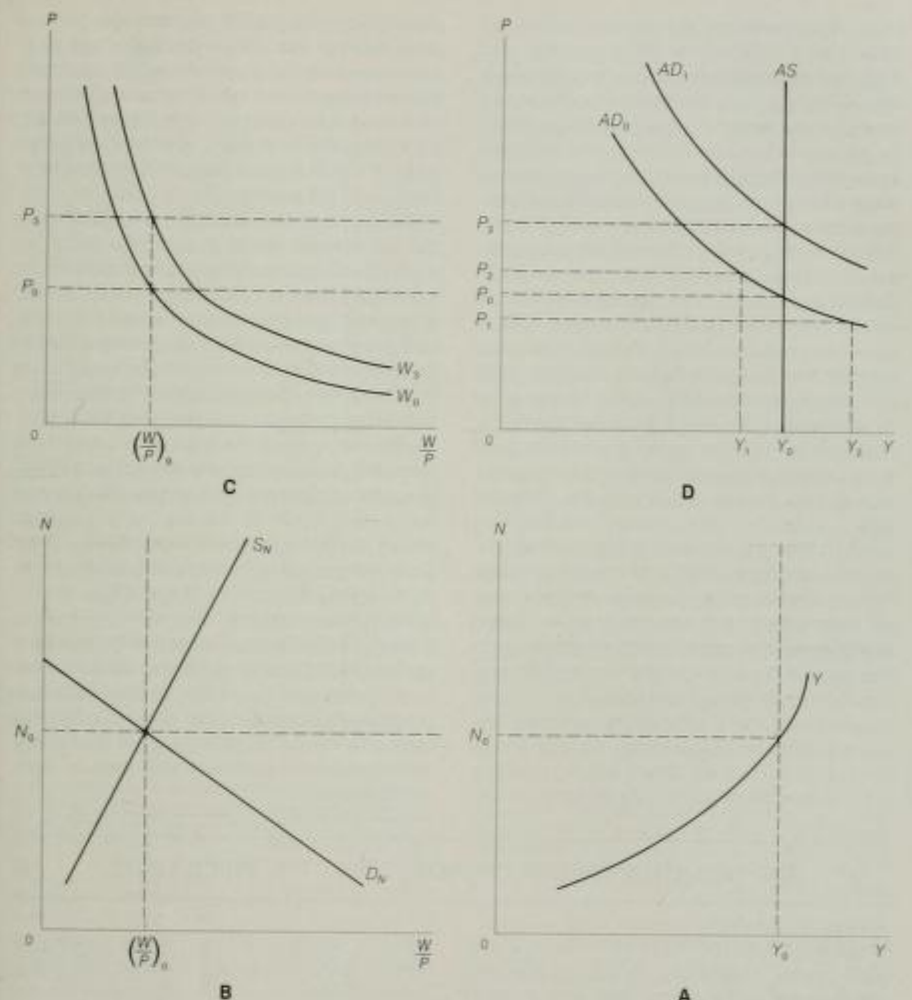


FIGURE 14-1
The classical model without saving and investment

amount necessary to establish equilibrium in the labor market. To review the analysis of the preceding chapter, at a real wage rate greater than $(W/P)_0$, there is an excess supply of labor. To achieve equilibrium, the real wage must fall and, on the assumption of a flexible money wage rate, this would be accomplished by the appropriate decline in the money wage rate.

Once W has fallen by the amount required to reduce W/P to $(W/P)_0$, equilibrium is reached. If we had assumed instead an initial real wage below $(W/P)_0$, the adjustment process would be similar. However, in this case there would be an excess demand for labor, and competition among employers would bid up the money wage rate by the amount necessary to raise the real wage to its equilibrium level.

Given such flexibility of the money wage rate, a displacement of the real wage from the equilibrium wage of $(W/P)_0$ that results from a change in P will be followed by a return to that real wage, because W will change as needed to offset that change in P. And as long as W/P is at $(W/P)_0$ for any value of P, Y also is at Y_0 for any value of P. Recall from the preceding chapter that this is the basis on which the perfectly inelastic aggregate supply curve in Part D of the diagram was derived for the classical case.

Recall also that the equilibrium level of employment found at $(W/P)_0$ is the full employment level; that is, at this level all those who are able, willing, and seeking to work at prevailing wage rates are able to find employment. This does not mean that there is neither any unemployment nor unfilled vacancies. At any time there will be some workers who are between jobs for various reasons, but unemployment of this kind is consistent with a condition of full employment as long as persons so unemployed are able to find positions in a reasonable period of time. Although there is some difference among economists over the appropriate percentage, most agree that unemployment of around 3 to 4 percent is accounted for by frictions of the kind noted, so that unemployment of 3 to 4 percent may be described as a situation of full employment.

Because any level of employment other than the full-employment level is a disequilibrium level, a familiar proposition of classical theory is that the equilibrium position in the labor market is necessarily one of full employment. Whatever unemployment, apart from frictional unemployment, persists in the face of this equilibrium must be voluntary unemployment. For example, with the equilibrium real wage established at $(W/P)_0$, any persons who are unemployed are considered to be voluntarily unemployed if they are seeking work but will accept work only at a money wage that, at the existing price level, means a real wage greater than $(W/P)_0$. They are seeking a real wage that is inconsistent with the marginal productivity of their labor, and in effect they are "pricing" themselves out of the market. Given the assumptions of the classical system, there is no barrier in that system to full employment or to the elimination of involuntary unemployment as long as labor is willing to reduce its money wage as required to produce that real wage at which all who wish to work will be hired.

THE QUANTITY THEORY OF MONEY AND THE PRICE LEVEL

As shown in Part D of Figure 14-1, the classical model yields a perfectly inelastic aggregate supply curve at the output level that is consistent with full employment. While this curve specifies what the output level must be in this model, it does not specify which of the various price levels along this supply curve will be the prevailing price level. To do this, classical theory relied on the quantity theory of money, a theory which says that it is the quantity of money in the hands of the public that determines how high or low the price level will be.

In dealing with money, we will accept the ordinary definition which states that money is anything that is generally accepted by the public in payment for goods, services, and other valuable assets and in the discharge of debts. Strictly speaking, only currency, coin, and bank

demand deposits qualify under this definition. Time and savings deposits, savings and loan shares, U.S. Savings Bonds, U.S. Treasury bills, and other federal government obligations near maturity come close to being money in this sense, for they may usually be converted into money quickly and with practically no loss of value. However, since these assets cannot generally be used to make payment until converted into coin, currency, or demand deposits, they do not fully qualify as money and are referred to instead as "near-money." If we choose to define money narrowly as anything that is generally acceptable as a means of payment, the total money supply at any point in time equals the sum of currency (which hereafter will be understood to include coin) and demand deposits that are held by the public.[4] Unless otherwise stated, each reference to the stock of money in this book is to money thus narrowly defined.

The Quantity Theory as a Theory of the Price Level

Although the public may choose to use currency and demand deposits as a store of value, i.e., as a form in which wealth can be held, the classical view was that the public uses other assets like interest-bearing savings deposits and securities as a store of value. It was maintained that the assets that make up money narrowly defined were used almost exclusively as a medium of exchange. The function of these assets or money in their role as a medium of exchange is simply to overcome the difficulties unavoidable in barter exchange. But even with money's role thus limited, a question remains: Does a bushel of wheat exchange for \$2 and a ton of coal for \$20, or does a bushel of wheat exchange for \$5 and a ton of coal for \$50? The answer given by classical theory is that the absolute level of prices, \$2 or \$5 for wheat and \$20 or \$50 for coal, depends on the quantity of money in the economy. This relationship in which the price level is made a function of the money supply is known as the *quantity theory of money*. Furthermore, the relationship between changes in the money supply and changes in the price level was held to be strictly proportional. This conclusion depended on several assumptions that may most simply be brought out by examining the identity $MV = PY$, in which M is the supply of money, V is its velocity, or the number of times it turns over per period in the purchase of final output Y, and P is the price level of this output.[5]

It should be emphasized that $MV = PY$ is an identity and as such stands completely apart

[4] The monthly *Federal Reserve Bulletin* includes a table headed "Measures of the Money Stock" that gives figures for five of the many possible definitions of the money stock: M_1, or currency and demand deposits held by the public, which is the narrowest definition limited to means of payment; M_2, or M_1 plus time deposits at commercial banks other than large Certificates of Deposit; M_3, or M_2 plus deposits at nonbank thrift institutions; M_4, or M_2 plus large Certificates of Deposit; and M_5, or M_3 plus large Certificates of Deposit.

At the time of writing, the only obligation of financial institutions counted in M_1 were demand deposit liabilities or those subject to transfer by ordinary check. However, since March 1976 thrift institutions and commercial banks in all of New England have had the power to offer NOW (negotiable order of withdrawal) accounts. These are accounts on which the holders receive interest and with which they can make payment in a way similar to drawing a check on a demand deposit. The amount of these accounts has understandably grown substantially and their exclusion from M_1 means that that figure understates the total means of payment. The appearance of these accounts has also raised difficult problems of competitive advantages and disadvantages among financial institutions which may be met in the not distant future by permitting commercial banks to pay interest on ordinary demand or checking-deposit balances.

It is convenient to use the letter M for the money stock as is done in this and following chapters and to use it also for imports as was done in Chapters 2 and 7 and as will be done in following chapters. In almost all cases it will be clear from the context which of the variables M refers to; where there may be any doubt, a specific statement will remove the doubt.

[5] M is a stock variable and Y a flow variable. If Y is defined for one quarter, M is the average stock of money in the economy during that quarter, V is the number of times that average stock of money is used to purchase final output during that quarter, and P is the average price level of output for that quarter.

from the quantity theory or any other theory. $MV \equiv PY$ says simply that the quantity of money multiplied by the number of times each unit of money on the average is spent for final output in any time period equals the quantity of final goods and services sold during that period multiplied by the price level of those goods and services. If Y is the physical volume of goods and services represented by any period's GNP, P is the price level of these goods, and V is the number of times the money supply is used to purchase goods whose value is PY, the familiar GNP identity (neglecting net exports), GNP $\equiv$ C + I + G, may be expanded to read $MV \equiv \text{GNP} \equiv C + I + G \equiv PY$. Each part of this expression set off by an identity sign is identical in value to every other part; each part is merely a different way of describing the same dollar amount.

The $MV \equiv PY$ identity is converted into the quantity theory of money under the assumptions that Y and V are constant or stable in the short run and that P is passive. With Y and V constant, the assumption that P is passive means that P depends on changes in M rather than that changes in M depend on changes in P. Given these assumptions, any short-run increase (or decrease) in M must lead to a proportional rise (or fall) in P.[6] Without these assumptions, however, it is equally inevitable that any increase (or decrease) in M will *not* lead to a proportional rise (or fall) in P (barring the unlikely case in which changes in V and Y are exactly offsetting).

The classical view that the level of output is stable in the short run simply reflects the fact that the fully employed labor force works with a fixed stock of capital and given production techniques in the short run. In terms of Figure 14-1, the production function in Part A can shift outward or to the right with growth in the capital stock and technological advances, resulting in a rise in the level of output consistent with full employment which appears in Part D as a rightward shift in the perfectly inelastic aggregate supply curve. However, these changes occur only gradually over the long run. The labor supply curve in Part B can shift to the left with the same result, but this too is a change that occurs gradually with the growth in population over the long run. We will trace changes of these kinds through figures like Figure 14-1 later in this chapter. Finally, short-run variations in output could appear as a result of departures from the normal position of a fully employed labor force, but such departures were regarded as infrequent and subject to prompt correction in a system of competitive markets. Thus, given the assumption that full employment of the labor force is normal, the assumption of a stable level of output for any short-run period follows logically.

The classical view that the velocity or turnover rate of money is constant is based on the argument that the institutional, structural, and customary conditions that determine velocity usually change very gradually. Among these conditions are the frequency with which economic units receive and make payments, the regularity of these receipts and payments, and the portion of such receipts and payments that are on a money or barter basis.[7] Though these and all other conditions affecting the size of V

[6] For example, begin with $100 \times 4 = 2 \times 200$. Increase M by 10 percent to 110. On the assumptions of the constancy of V and Y, we have $110 \times 4 = 2.2 \times 200$, or a 10-percent increase in P, a rise proportional with the increase in M. Or, alternatively, rewrite the identity as $M(V/Y) \equiv P$, in which the assumed constancy of V and Y makes V/Y a positive proportionality constant, here equal to 4/200, or 0.02. Then, whatever the value for M, P is always 0.02 times that value.

[7] Even though no one may choose to hold idle money, everyone holds some currency or demand deposits to even out the difference between receipts and payments. For example, if an employee is paid \$200 every other Friday, he does not typically spend the whole \$200 on payday (and end up "broke" for the next thirteen days). If he spends the \$200 evenly over the two-week period, his average cash balance will turn out to be \$100. Since his biweekly spending is \$200, this \$100 average balance has a V of 2 biweekly; since his annual spending is \$5,200, this \$100 average balance has a V of 52 annually. If he were instead paid \$100 every Friday, by the same line of argument his average cash balance would be \$50. This

are subject to change, the quantity theory asserts the gradualness of such change in support of its conclusion that V is constant in the short run.[8]

The Quantity Theory as a Theory of Aggregate Demand

Whereas the simple Keynesian model is unable to say anything about the level of output without a theory of aggregate demand, the simple classical model is able to establish the level of output and apparently also the price level of output without such a theory. The output level is determined to be that output which is produced by a fully employed labor force and the price level at which that output sells is determined by the quantity of money. Although the classical theorists thus seemed to make do without explicitly bringing in aggregate demand, implicit in their quantity theory of money is a theory of aggregate demand. This theory of aggregate demand may be described with an appropriate aggregate demand curve. Then the combination of that aggregate demand curve and the aggregate supply curve which is the heart of the classical model gives us a simple aggregate demand—aggregate supply classical model which can be compared with the simple Keynesian model developed in Part Two. The theory of aggregate demand in this classical model is crude, and the aggregate demand curve that describes it is easily derived.

To begin with, classical theory held that, whatever the size of the money supply, the full amount was in active circulation or was all in use as a medium of exchange under normal conditions. This is the assertion that the public does not hold money as a store of value. The absence of such idle money balances is a critical element in the rigid quantity theory, and, in the view of the classical theorists, the absence of such holdings was what was to be expected on the basis of rational human behavior. They could see no reason why people should choose to hold any portion of their money receipts in idle money form. As people came into the possession of money, there was in this view only one disposition for it: spending. Spending was either for consumption or for capital goods. As we will see, the act of saving, or not spending for consumption goods, was automatically transformed into an act of spending for capital goods. Money that was held back from consumption spending would be loaned to firms that would, in turn, spend the money, dollar for dollar, for capital goods. Thus, although persons do save, classical theory held that they would not hold any amounts saved in the form of money and that therefore no money would escape from active circulation.

As long as money was used exclusively as a medium of exchange and thus remained completely in active circulation, the velocity of money would remain stable. With V known and stable, MV or total spending on goods and services is known as soon as M is known. This brings us to the concept of aggregate demand as it appears in the simple classical model of Figure 14-1. It is portrayed graphically by a curve like AD_0 in Part D. That curve shows the various quantities of output that can be purchased at various price levels with a certain fixed amount of total spending. As V is stable in the simple classical model, what that amount of

$50 average balance has a V of 4 biweekly and a V of 104 annually. Generalizing for the economy as a whole, with a given supply of money, a change in which everyone were paid half as much twice as often would mean the existing supply of money could handle a much greater volume of final purchases. This rise in V with constant M would mean a rise in PY proportional with the rise in V.

This illustration covers frequency of receipts, only one of the many conditions that were believed to change very slowly and thus to make for stability in V. For a discussion of the determinants of V, see J. M. Culbertson, *Money and Banking*, 2nd edition, McGraw-Hill, 1977, pp. 125–32, and G. Garvy and M. R. Blyn, *The Velocity of Money*, Federal Reserve Bank of New York, 1969, Ch. 6.

[8]Even in its crude form, the quantity theory did not argue that short-run V and Y were as stable or P as passive as is here assumed. We will retain these assumptions in the extreme form in which they have been given so that we may proceed with the construction of the simplified classical model.

spending will be in any time period depends entirely on M, the average stock of money held by the public during that time period. For each stock of M, there is a different aggregate demand curve: the greater the stock of money, the farther from the origin is the corresponding aggregate demand curve.

For any given AD curve, every combination of P and Y along that curve will be a combination whose product equals the same total spending, a property which tells us that the AD curve in question is of unitary elasticity throughout.[9] However, we have seen that the classical theory makes Y independent of total spending: Y is whatever a fully employed labor force is able to produce. It is shown in Part D by the position of the perfectly inelastic AS curve. Accordingly, the intersection of the AD curve with the AS curve does no more than determine the price level. Thus, for AD_0, the equilibrium price level must be P_0. Any departure from this price level sets into motion forces that return P to P_0.

If, for example, P were P_2 instead of P_0, the level of total spending indicated by AD_0 would be able to purchase output of only Y_1, while the amount of output supplied would be unchanged at Y_0. This is a case in which $AS > AD$. Does this mean that producers will adjust by cutting back from the Y_0 to the Y_1 level of output and by laying off workers? Not at all—production will remain at Y_0. Given the money wage rate of W_0, producers find it consistent with the requirement for profit maximization—$W = P \times MPP$—to supply Y_0 of output at a price level of P_0. At the higher price level of P_2, they cannot sell all they can produce at full employment, but the result is not a reduction in production. It is the price level that adjusts downward as producers compete for customers, and once the price level has fallen to P_0, buyers can again purchase with the same total spending the full employment output that producers are prepared to supply at this price.

If we choose a disequilibrium level for P below rather than above P_0, an upward adjustment occurs. At the lower price level of P_1, the total spending represented by AD_0 calls for output of Y_2, which exceeds the output produced at full employment. The result in this case must be a rise in P. Competition among buyers who, per time period, spend the unchanged sum represented by AD_0 will force P up until equilibrium is restored with P at P_0.

To sum up, on the assumption of competitive conditions in the markets for output as well as the markets for labor, the classical model assures that the equilibrium level of output will be the full employment level or Y_0 as shown in the illustration by the intersection of the AS and AD_0 curves. A deficiency of aggregate demand that would appear if P were above the level given by the intersection or an excess of aggregate demand that would appear if P were below the level given by the intersection is promptly removed as competitive forces drive P to the level at which $AS = AD$.

CLASSICAL MODEL WITHOUT SAVING AND INVESTMENT

The interconnected parts of Figure 14-1 enable us to identify the full set of equilibrium values for this simple classical system. Thus, these values are N_0, $(W/P)_0$, Y_0, W_0, and P_0. Barring

[9] Like the W curves in Part C, the AD curves in Part D are rectangular hyperbolas. Each AD curve corresponds to a particular value of MV, and MV is by definition equal to PY. Given the identity between MV and PY, every rectangle found by running perpendiculars from any point on a given AD curve to the axes must be equal in area. As the AD curve is such that PY is equal for every pair of P and Y values along it, price elasticity of demand is unity at every level of P.

any shift in the production function or the supply curve of labor or any change in the money supply or its velocity, the indicated set of equilibrium values will remain unchanged period after period. In practice, of course, these elements will change over time, but for each change, under classical assumptions, new equilibrium values will be established for the variables of the system. Tracing through several such changes will illustrate the mechanics of the system.

Effects of a Change in the Supply of Money

Consider first the case of an increase in the money supply from, say, M_0 to M_1, which produces the shift of the *AD* curve from AD_0 to AD_1 in Part D of Figure 14-1. The increase in *M* (with constant *V*) means an increase in total spending per time period of $V(M_1 - M_0)$ and a rise in the price level from P_0 to P_3. If the money wage does not rise proportionally with this rise in the price level, the real wage will fall, causing employers to try to expand output by hiring more workers, as a higher price for output without a higher money wage rate means greater profits with greater output. But a real wage below $(W/P)_0$ means that the quantity of labor available is too small to produce output Y_0, let alone to expand output beyond this. Competition among employers for workers will then force the money wage up to W_3, or proportionally with the price level, leaving the equilibrium real wage unchanged at $(W/P)_0$ and output unchanged at Y_0. The net result of the expansion of the money supply is a proportional rise in the price level and in the money wage but no change in employment or output; the equilibrium values are N_0, $(W/P)_0$, Y_0, P_3, and W_3.

This, of course, is just what we should expect on quantity theory reasoning. The level of output is determined by the aggregate supply curve; the money supply only sets the price level for this output. Increasing or decreasing the money supply will cause the price level of output to rise or fall proportionally, but the level of output itself will remain unchanged at the full employment level. Any change in the money supply that is accompanied by a change in the velocity of money will break the proportional relationship between *M* and *P* but will still leave the level of output and employment unaffected by changes in either *M* or *V*.

Effects of a Change in the Supply of Labor

Now let us imagine an increase in the labor supply, as shown by the shift from S_{N_0} to S_{N_1} in Part B of Figure 14-2. With no shift in the production function and so no shift in the curve of the MPP of labor, any increase in employment will lower the MPP of labor. The full employment equilibrium previously was at N_0 with the real wage of $(W/P)_0$. The new full employment equilibrium is at N_1, and to achieve this equilibrium requires a decline in the real wage from $(W/P)_0$ to $(W/P)_1$. At this real wage and level of employment, output is seen to be Y_1 in Part A and the *AS* curve in Part D is AS_1, a shift to the right from its original position at AS_0.

What is required to bring about an adjustment to the new full employment equilibrium following the increase in the supply of labor? With *MV* unchanged, *AD* in Part D remains at its original position of AD_0. Therefore, the sale of the larger output produced by the larger number of workers employed at the new full employment position can only occur at the lower price level of P_1. One thing then required is the indicated decline in *P*. Producers will find it to their advantage to produce this larger output and sell it at the lower price level if the money wage rate declines enough. What specifically is required is that *W* fall in Part C from W_0 to W_1 as it is only with *W* of W_1 that *P* of P_1 will yield $(W/P)_1$, the real wage consistent with full employment after the increase in the labor supply.

This decline in *W* must clearly be greater than the decline in *P* in order to bring about the required decline in *W/P*. To see this graphically, suppose for the moment that *W* declines only

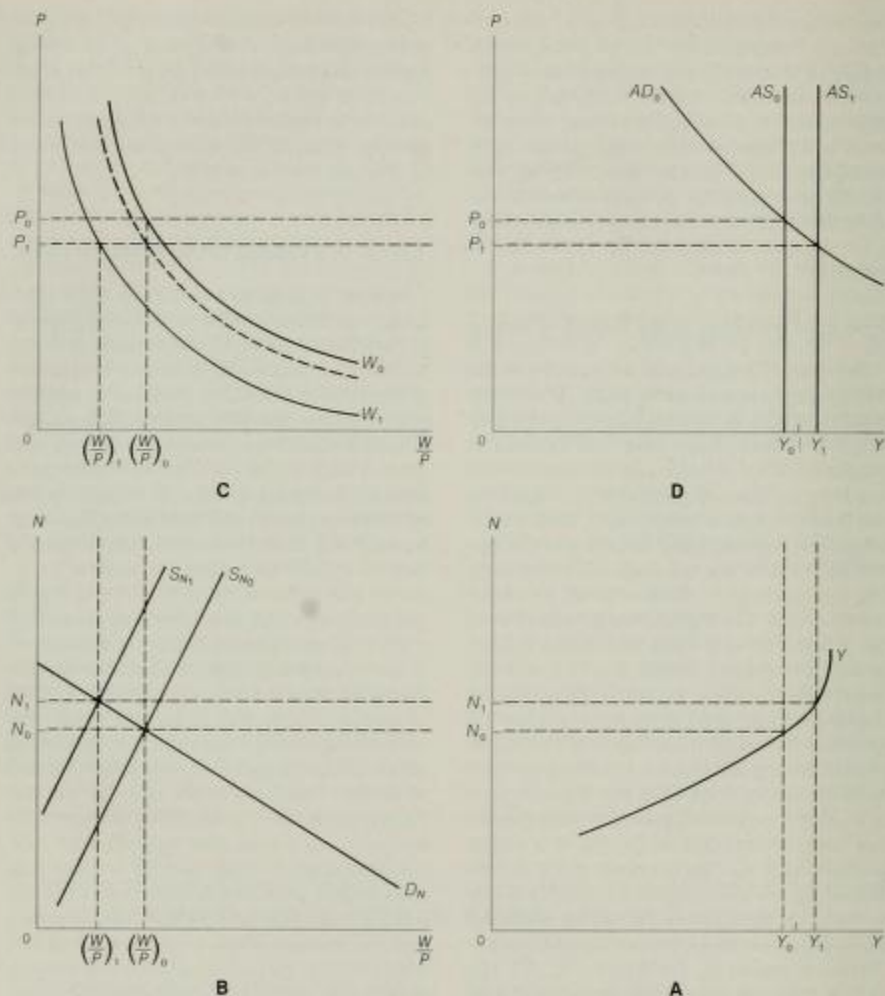


FIGURE 14-2
The classical model: effects of a change in the supply of labor

by the amount indicated by the broken line between the W_0 and W_1 curves. This decline would be proportional with the decline in P as a horizontal line from P_1 intersects the broken W curve at the same W/P at which a horizontal line from P_0 intersects the W_0 curve. But if W

fell only that far, there would be no decline in the real wage and therefore no incentive for employers to hire more labor and expand output in the first place. Furthermore, if W fell proportionally with P and thus maintained Y at its original level of Y_0, there could be no basis for P to fall in the first place. A decline in P below P_0 with AS remaining at AS_0 and Y therefore remaining at Y_0 would create excess demand, $AD > AS$, the result of which would be a return of the price level to P_0. However, given the classical theory's assumption of downward wage flexibility, W will fall any time there is less than full employment, and, given the increase in the supply of labor and the other assumptions, a fall in W to W_1 will occur to restore the system to full employment. With W down to W_1, a new full employment equilibrium with the following values is established: N_1, $(W/P)_1$, Y_1, W_1, and P_1.

A numerical "before" and "after" example may clarify the adjustments involved. The first row of the table above gives the equilibrium values for N, Y, W, P, and W/P when labor demand is D_N, labor supply is S_{N_0} (Figure 14-2), the money stock is $75, and velocity is 4. The second row gives the new equilibrium values after the shift in S_N to S_{N_1}. Full employment now calls for N of 150, at which the MPP of labor and therefore W/P is 1.66. With N of 150, the full employment Y is 400. If Y is 400, P must be 0.75. With MV and AD unchanged at 300, 400 of output can be sold only at P of 0.75. Finally, with P of 0.75, W must adjust downward to the degree required to achieve the new real wage of 1.66, if there is to be full employment.[10]

	N	Y	W	P	$MPP = (W/P)$	M	V
Original equilibrium (subscripts 0)	100	300	$2.00	1.00	2.00	$75	4
New equilibrium (subscripts 1)	150	400	1.25	0.75	1.66	75	4

Given an increase in the labor supply, the crucial element of the process by which the system moves to its new equilibrium position is the adjustments that occur in the money wage and the price level. Whether unemployment results from an increase in the labor supply or for other reasons, flexibility of the money wage and price level is indispensable to the correction of unemployment. As long as the money wage responds to unemployment and as long as the price level responds to changes in output, full employment can always be regained according to this simple classical model.

Effects of a Change in the Demand for Labor

Growth in the capital stock or technological advances will cause the production function to shift outward over time, as shown by the movement from Y to Y' in Part A of Figure 14-3. At each possible level of employment, the MPP of labor is now greater than it was, since at each level of N the slope of Y' exceeds the slope of Y. This is reflected in Part B of Figure 14-3 as an upward or rightward shift in the demand curve for labor, indicating that it is now profitable for employers to hire more labor at each possible real wage. The equilibrium real wage

[10] On the assumption of profit maximization, employers will not expand employment unless greater profits are expected from the sale of the higher level of output. In this case, there will be greater profits, as may be seen from the figures. At the original equilibrium, labor's share of the real output of 300 is $N \times$ MPP, or 100×2, or 200. The remainder, $Y - (N \times \text{MPP})$, or 100, may be called the "profit share." At the new equilibrium, labor's share of the real output of 400 is 150×1.66, or 250, leaving 150 as the "profit share," an increase in profits of 50. In dollar terms, the flow of income at the original equilibrium is $300, or 300×1, divided into $200 for labor and $100 for profits. At the new equilibrium, it is the same $300, or 400×0.75, now divided into $187.50, or $150 \times \$1.25$ for labor and $112.50 for profits. Though labor's share is decreased in money terms from $200 to $187.50, the $187.50 adjusted for the fall in P from 1 to 0.75 is equal to $250 in "base period" prices. Similarly, the profits of $112.50 are equal to $150 in "base period" prices.

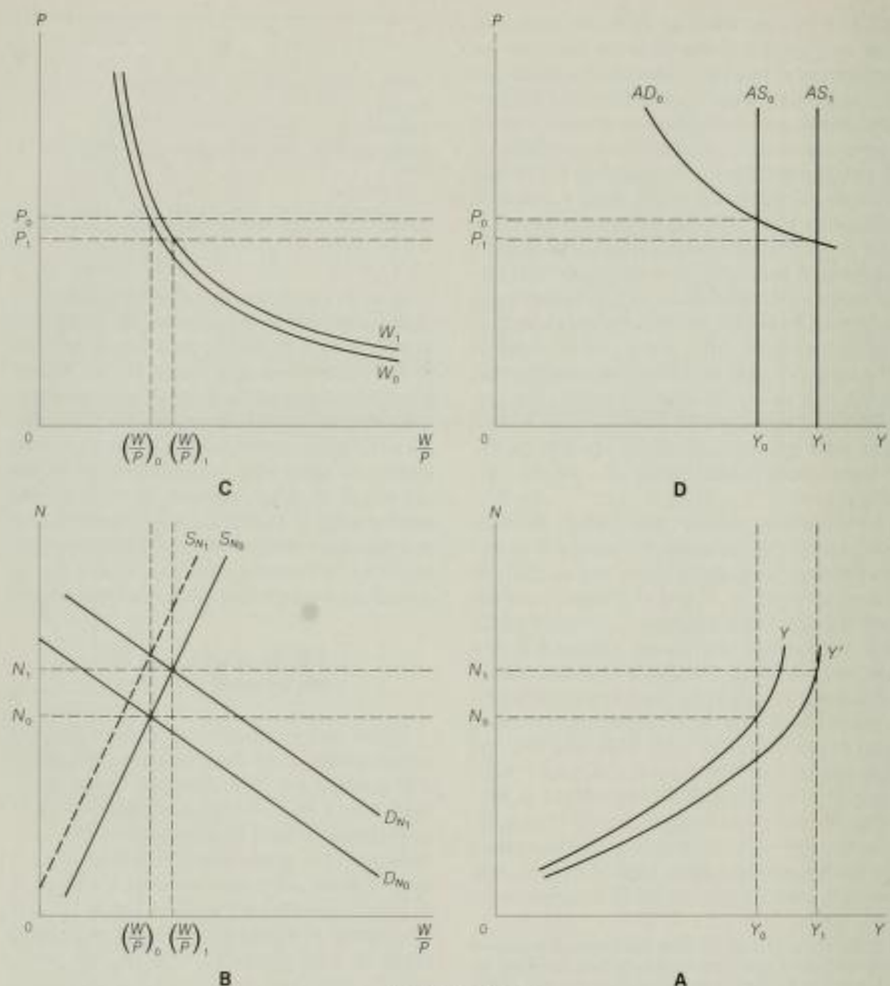


FIGURE 14-3
The classical model: effects of a change in the demand for labor

rises from $(W/P)_0$ to $(W/P)_1$, employment rises from N_0 to N_1; AS shifts from AS_0 to AS_1 and output rises from Y_0 to Y_1. With no change in the money supply, the greater output requires a fall in the price level from P_0 to P_1. At the new equilibrium real wage of $(W/P)_1$, price level P_1 calls for money wage W_1. In the present case, the rise in the real wage necessary to reestablish

equilibrium is produced by a fall in P and a rise in W.

The first row of the following table repeats the set of figures previously used to describe the original equilibrium values; the second row gives a set of figures that describes the new equilibrium.

	N	Y	W	P	MPP = (W/P)	M	V
Original equilibrium (subscripts 0)	100	300	\$2.00	1.00	2.00	\$75	4
New equilibrium (subscripts 1)	105	333	2.25	0.90	2.50	75	4

In passing, we may note several basic propositions in economics that are clearly brought out by the present analysis. For one, the gradual rise in the real wage, or "standard of living" of labor, is primarily the result of the gradual outward shift in the production function, which is largely attributable to technological progress and a growing stock of capital. If, over the same period in which these developments raised the schedule of marginal productivity of labor from D_{N_0} to D_{N_1}, the supply curve of labor had also moved from S_{N_0} to S_{N_1}, the number of workers employed would have risen, but the real wage would have remained the same. The actual gradual rise in the real wage experienced over the long run in Western economies has resulted primarily from the fact that the growth in capital and the rate of technological advance have exceeded the rate of growth in the labor force.

A second proposition brought out by this analysis is that the long-run growth of output (whether produced as here by an outward shift in the production function, or as in the previous case by a shift to the left in the labor supply curve, or as in practice by both) leads to a falling price level unless accompanied by an expansion in the money supply. Although an expansion of output with no rise in the money supply will in practice cause V to rise, V cannot rise without limit. Therefore, as output expands in the long run, M must expand to avoid what otherwise must be a gradually falling P. Although these propositions have been brought out by the classical model, they are accepted in principle by most economists today.

Effects of a Rigid Money Wage

From an initial equilibrium with a given AD curve, an increase in the supply of labor calls for a fall in the price level and a larger fall in the money wage rate to establish a new full-employment equilibrium at a lower real wage. Classical theory assumes perfectly competitive markets, and an excess supply of labor in such markets will automatically depress the money wage. If we now drop the assumption of perfect competition in the labor market, the results may be different. Consider, for example, the imperfection of competition that results when workers are organized into labor unions. There will be no barrier to a rise in the money wage when excess demand for labor appears, but there will now be a barrier to a fall in the money wage when excess supply appears. In other words, the money wage is flexible upward but may be rigid downward. Furthermore, in an imperfect market the money wage may be forced up even though there is no excess demand for labor. To illustrate, let us begin with a full-employment equilibrium position for the economy and observe the effects that follow when the money wage is arbitrarily pushed up, say by union pressure.

In Figure 14-4 there is full-employment equilibrium with a real wage of $(W/P)_0$ and values for other variables indicated by subscript 0. Suppose now that the money wage is forced up from W_0 to W_1. If the price level were to remain at P_0, the real wage would rise proportionally to the rise in the money wage. But, with the given AD, the price level must rise; for in the absence of a rise in P there is a rise in the real wage, which means a decrease in the amount of Y supplied, and with aggregate demand given by AD_0, a smaller Y is accompanied by a higher P. While P must therefore rise, it cannot, however, rise as far as W, for if it did there would be

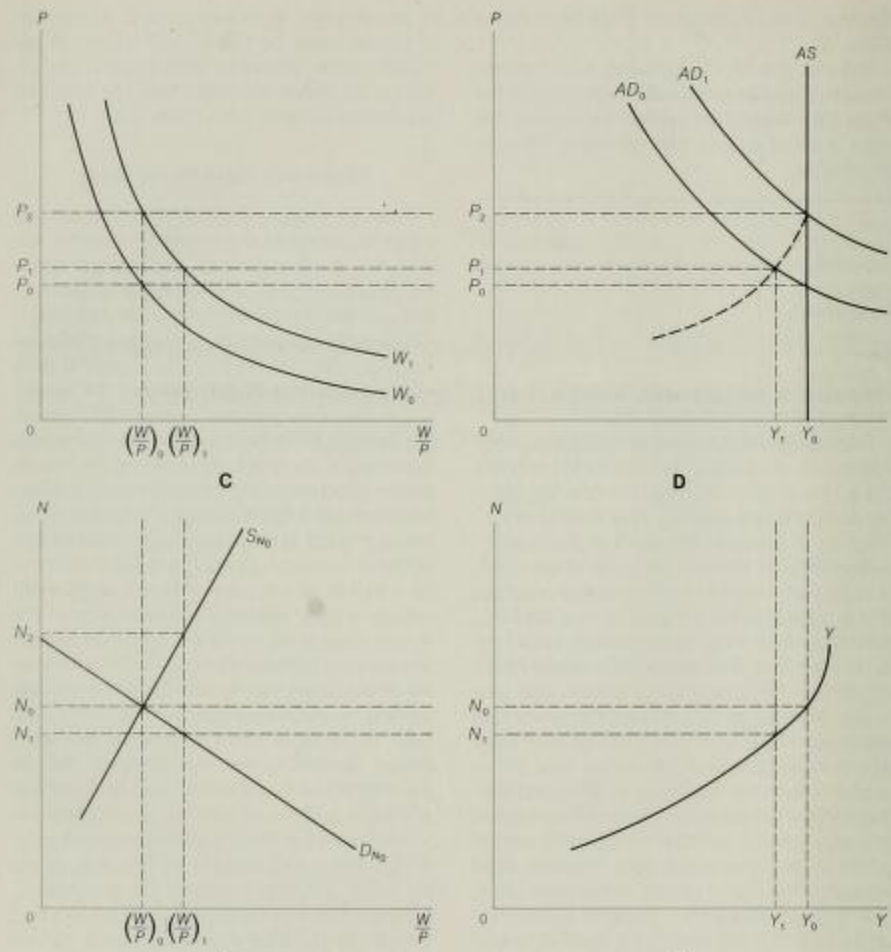


FIGURE 14-4
The classical model: effects of a rigid money wage

no change in the real wage and so no change in output. The original output cannot all be sold with unchanged aggregate demand at a higher P, so P must rise in proportion to the fall in output.

The process by which a new equilibrium is

reached is one in which P, Y, and N must all adjust to the fixed money wage, W_1. The new equilibrium values for P, Y, and N are designated by the subscript 1. As compared to the initial equilibrium, there are now a higher real wage, a lower output level, and a higher price level. The higher real wage, which was artificially brought about by forcing up the money wage, forces employment down from N_0 to N_1. Since the amount of labor supplied is greater with a higher real wage, the amount of unemployment is not merely the difference between N_1 and N_0 but the larger difference between N_1 and N_2. With the higher real wage, those workers fortunate enough to keep their jobs are, of course, better off than they were before the rise in the money wage.

Although we will not go into this part of the model, note that the arbitrary increase in W from W_0 to W_1 makes the AS curve slope downward as shown by the dashed segment of AS. An AS curve which is perfectly inelastic throughout can only be obtained on the assumption of a flexible money wage rate. Drop that assumption as we have temporarily done here, and the effect on the AS curve is as shown, the basis for which was examined in detail in the preceding chapter.

The table below provides a numerical example, similar to those given earlier, which illustrates the results that follow from the introduction of a rigid money wage into the present model. Since the new equilibrium is not one of full employment, the last three columns have been added to show the resulting unemployment. (S is labor supplied; D is labor demanded; U is labor unemployed.) With full employment equilibrium defined by a real wage of 2, we can see from the figures that unemployment must result from the arbitrary raising of the money wage from \$2.00 to \$2.40.

As long as the money wage is arbitrarily held above the level consistent with full employment, we have an *equilibrium situation with unemployment*. Although we have noted several times before that classical theory denied this possibility of equilibrium with unemployment, the denial was made only on the assumption that we were dealing with an economy in which the money wage was flexible. An equilibrium with less than full employment is therefore entirely consistent with classical theory when that theory is stripped of the assumption of flexible wage rates, an assumption indispensable to its full-employment conclusion.

In the *General Theory*, Keynes replaced the classical assumption of a flexible money wage with that of a rigid money wage, an assumption certainly more closely in agreement with the facts of observation. In so doing, Keynes could easily enough show that equilibrium with unemployment is possible. Though a great deal more is involved, what should be clear from analysis of the present case is that the corresponding change in assumption in the classical theory leads to the same possibility of equilibrium with unemployment reached by Keynes in the *General Theory*.

								Labor		
	N	Y	W	P	$MPP = W/P$	M	V	S	D	U
Original equilibrium (subscripts 0)	100	300	\$2.00	1.00	2.00	\$75	4	100	100	0
New equilibrium (subscripts 1)	90	273	2.40	1.10	2.18	75	4	102	90	12

Monetary Policy and Full Employment

In the classical scheme, if the money wage is held artificially above the level necessary for full employment, an appropriate expansion of the money supply may be an antidote. According to the quantity theory, an increase in M, with V and Y stable, will raise P proportionally. With a rigid money wage, the rise in P reduces the real wage and provides the profit incentive to employers to expand employment and output toward the full-employment level. There is, therefore, some appropriate expansion in the money supply that is sufficient first to raise P to the level that reduces the real wage, W/P, to the full-employment equilibrium level and second to provide the increase in AD needed to purchase the full employment output at that price level.

In terms of Part C of Figure 14-4, to achieve the full-employment real wage of $(W/P)_0$ with the money wage inflexible at W_1 requires a price level of P_2, since at that level W_1/P_2 equals W_0/P_0, or $(W/P)_0$. With real wage of $(W/P)_0$, output is Y_0. Therefore, in Part D, MV must be increased to equality with P_2Y_0 to generate AD adequate to purchase full-employment output Y_0 at price level P_2.[11] The AD curve must shift to AD_1.

[11] With V constant, $\Delta M = M_1V - M_0V = P_2Y_0 - P_1Y_1$.

The previous numerical example may be modified to show how an appropriately expansionary monetary policy may offset the effect of a rigid money wage. The first two rows of the table below are the same as before—the first describes the initial full-employment equilibrium, the second the less-than-full-employment equilibrium that results from a money wage artificially pushed up. The third row shows the return to a full-employment equilibrium that results from the appropriate expansion of M.

Note that the strict quantity theory does not hold in this situation, because part of the increase in AD created by the expansion of M is absorbed by the expansion of Y that is called forth by the fall in the real wage. M rises from \$75 to \$90, or by 20 percent, so that the strict quantity theory would indicate a 20-percent rise in P or from 1.10 to 1.32. P actually rises from 1.10 to 1.20 or by less than 10 percent as Y rises from 273 to 300.

Thus, it would seem that monetary policy provides the solution to unemployment created by a rigid money wage. But it is equally apparent from the crude model before us that this method of securing full employment in the face of an artificially high wage rate works only as long as the increase in M is not offset by a decrease in V. Aggregate demand must increase with the increase in M. Classical theory saw no "leakage" between an increase in M and an

								Labor		
	N	Y	W	P	$MPP = W/P$	M	V	S	D	U
Original equilibrium (subscripts 0)	100	300	\$2.00	1.00	2.00	\$75	4	100	100	0
Second equilibrium (subscripts 1)	90	273	2.40	1.10	2.18	75	4	102	90	12
Third equilibrium (N_0, Y_0, W_1, P_2)	100	300	2.40	1.20	2.00	90	4	100	100	0

increase in aggregate demand. We can begin to see why monetary policy was *the* policy weapon of classical economists. When we return in the next chapter to Keynesian theory, we will see that this simple tie between changes in the money supply and changes in aggregate demand disappears. In Keynesian theory, aggregate demand cannot be so simply increased or decreased by expansion or contraction of the money supply.

CLASSICAL MODEL WITH SAVING AND INVESTMENT

Although formally correct, the classical model we have been discussing is oversimplified, because it fails to break aggregate demand down into demand for consumption goods and demand for capital goods. This means that it does not recognize the processes of saving and investment.

Saving and Investment

We must now recognize that not every dollar of income earned in the course of production is spent for consumption goods; some part of this income is withheld from consumption, or saved. Clearly, unless there is a dollar of planned investment spending for every dollar of income saved, there is a deficiency of aggregate spending. Another part of classical theory provides the mechanism that presumably assures that planned saving will not exceed planned investment. This mechanism is the interest rate.

Classical theory treated saving as a direct function of the interest rate and investment as an inverse function, as illustrated in Figure 14-5. The investment curve is simply the curve of the marginal efficiency of investment (MEI), whose derivation was explained in some detail in Chapter 10. There it was a part of our development of Keynesian theory; here we see that this was actually a part of the classical theory taken over by Keynes. It is important to note, however, that although both theories show investment as an inverse function of the interest rate, this is not to say that both assign equal importance to the interest rate as an influence on investment spending. The whole question of the elasticity of the investment schedule is involved, a topic we discussed at length in Chapter 12. Keynes and most economists since Keynes have argued that the curve is relatively inelastic. If it is elastic, a relatively small change in the rate of interest will be sufficient to call forth a relatively large change in investment; relatively small changes in the interest rate will then be all that are required to keep planned saving and planned investment in balance as the saving and investment schedules shift. If it is inelastic, relatively large changes in the interest rate will be required for this purpose.[12] The question then arises as to whether the interest rate will fluctuate freely over the wider range necessary to equate saving and investment. To simplify the exposition of the classical system, let us assume here that the curve is indeed elastic, so that investment is relatively responsive to changes in the interest rate. Small changes will then keep saving and investment in balance.

The saving curve of Figure 14-5 is new. Here saving is made a direct function of the interest rate; in Keynesian theory, saving is a direct function of the level of income. In the Keynesian scheme, the interest rate may have an

[12] It is also conceivable that both the investment and saving curves are so inelastic that a shift to the right in the saving curve or a shift to the left in the investment curve or a combination of the two may result in an intersection of the two curves only at a negative interest rate. However low the interest rate might fall, it assuredly could not fall below zero. The result is an impasse at which the interest rate is completely powerless to equate saving and investment.

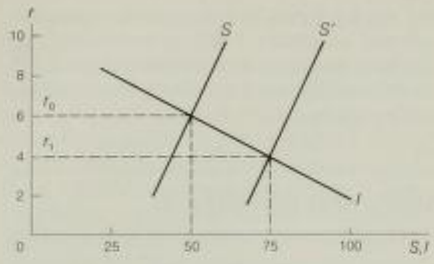


FIGURE 14-5
Classical equilibria between saving and investment

influence on saving, but it is of minor importance. In classical theory, the interest rate is all-important, and the level of income is of minor importance. Since the classical model argues that full employment is the normal state of affairs in the economy, the level of income is in effect ruled out as a variable in the short run, and so it is ruled out as an influence on the amount of saving. The problem in classical economics is to explain how saving will vary at the full-employment level of income, and the solution is provided by the interest rate. The higher the interest rate, the greater the amount of the full-employment income that is withheld from consumption or devoted to saving.

Given saving and investment curves such as S and I of Figure 14-5, competition between savers and investors would move the interest rate to the level that equated saving and investment. If the rate were above r_0, there would be more funds supplied by savers than demanded by investors, and the competition among savers to find investors would force the rate down. If the rate were below r_0, competition would force the rate up. When the rate is at r_0, equilibrium is established, with every dollar saved or withheld from consumption spending matched by a dollar borrowed and devoted to investment spending.

It is important to see that this transfer of money from savers to investors also involves a transfer of resources. The decision to save part of current income is a decision by income recipients not to exercise their claims to the full amount of output that results from their productive services. This releases resources from the production of consumption goods and makes them available for the production of capital goods. These resources will be fully absorbed in the production of capital goods only if investors choose to purchase exactly the amount of capital goods that can be produced by the resources released as a result of saving. This means that if the rate of interest were above r_0 and somehow stayed above r_0, unemployed resources would appear, for the excess of S over I at an interest rate above r_0 reflects, in real terms, an excess of resources released from the production of consumption goods over the amount absorbed in the production of capital goods. One of these resources is, of course, labor, and the excess of S over I also means that there is an excess of labor available over labor employed. In a word, there is unemployment. Thus, in the classical system, if the interest rate fails to equate saving and investment, it also fails in its assigned task of promptly reallocating the resources released from production of consumption goods to the production of capital goods, and unemployed resources are the result.

Changes in Saving and Investment

As long as the interest rate adjusts upward and downward to correct any disequilibrium, shifts in the saving and investment functions will lead to the establishment of new equilibrium positions. Suppose that income recipients become more thrifty; at each interest rate they choose to save a larger part of their current income. This appears in Figure 14-5 as a shift to the right from S to S' in the saving curve and a decrease in the interest rate from r_0 to the new equilibrium level r_1. A numerical example such

as those presented earlier brings out the effects of an increase in thrift in the classical system. The first row of the table below indicates the values of the variables at the original full-employment equilibrium. Full employment of the labor force is *N* of 100, and full-employment output is *Y* of 300. With the interest rate at r_0, say 6 percent, the real income of 300 is divided into 250 of consumption and 50 of saving; *r* of 6 percent also produces equilibrium with saving of 50 and investment of 50. If the saving curve now shifts to the right and the interest rate drops to r_1, say 4 percent, a new equilibrium is established at which saving of 75 (in real terms, *Y* of 300 less *C* of 225) equals investment of 75. With no shift in the production function or the supply of labor, full-employment output remains at 300. The only change is in the distribution of output between consumption goods and capital goods. Thus we see that the increased thrift of the public has produced a reallocation of resources—one away from the production of consumption goods and to the production of capital goods—but with the total production of goods unchanged at the full-employment level of 300.

If the saving function shifted in the opposite direction so that there were less saving at each interest rate, we would have a higher interest rate at which a smaller flow of saving would be equated with a smaller flow of investment but still with the flow of aggregate output unchanged from the full-employment level with which we started. The effects of shifts in the investment curve may be traced in the same way. Whatever the shifts in the saving and investment curves, however, the possibility of "oversaving" or "underconsumption" could not arise so long as the interest rate succeeded in balancing saving and investment.

Does the interest rate always adjust promptly as required to maintain equality between saving and investment? The Swedish economist Knut Wicksell (1851–1926) was one of the first to point out that there are conditions under which it would not. However, even in this event, the full-employment level of output would be maintained if prices and wages were sufficiently flexible. If the interest rate did not adjust promptly, there would be an excess of planned saving over planned investment or planned investment over planned saving. According to Keynesian theory, a fall or rise in income (output and employment) would be required at this point to bring saving and investment back into balance. This was not the case in classical theory, however. An excess of saving over investment would mean a deficiency of aggregate demand at the existing price level. This would lead to a deflation of prices and wages but would not interfere with the maintenance of the real wage consistent with full employment. An unchanged aggregate demand that had become deficient at the original price level would now be adequate to purchase the full-employment level of output at a lower price level. Conversely, an excess of planned investment over planned saving would mean an excess of aggregate demand at the existing price level. This would lead to an inflation of prices and wages but again would not interfere with the maintenance of the real wage consistent with full employment.

	N	*Y*	*C*	*I*	*W*	*P*	MPP = *W*/*P*	*M*	*V*
Original equilibrium (r = 6 percent)	100	300	250	50	\$2.00	1.00	2.00	\$75	4
New equilibrium (r = 4 percent)	100	300	225	75	2.00	1.00	2.00	75	4

SUMMARY STATEMENT

The purpose of this chapter has been to show, in terms of a simple model, how classical theory answered the fundamental questions of macroeconomics: What determines the levels of employment, output, consumption, saving, investment, prices, and wages? Our discussion may now be summarized in a list of the basic propositions that make up classical theory. Each of these propositions will be related to the graphic apparatus of Figure 14-6, which adds nothing new but brings together the saving–investment branch of the classical system with its other branches, presented in two steps in the preceding pages.

1. As shown in Part B, the supply, S_N, and demand, D_N, for labor are both functions of the real wage, W/P. Because of diminishing returns, the demand curve slopes downward to the right (i.e., more labor is hired only at a lower real wage). Because of the essential disagreeableness of work, the supply curve slopes upward to the right (i.e., more labor is offered only at a higher real wage). The intersection of the supply and demand curves thus determines both the real wage $(W/P)_0$ and the level of employment N_0, which is the full employment level.
2. With fixed techniques of production and fixed capital stock, output in the short run becomes a function of employment, as shown by the production function in Part A. With employment determined in Part B as N_0, output is determined in Part A as Y_0.
3. On the assumption of a flexible money wage, the aggregate supply curve, *AS*, in Part D is perfectly inelastic at Y_0, the full employment level of *Y*. The aggregate demand curve, *AD*, is determined by the stock of money, *M*, given the assumption that the velocity of money, *V*, is stable. The price level, P_0, is then determined by the intersection of the *AS* and *AD* curves in Part D. As *AS* is perfectly inelastic and as the position of *AD* depends on the size of the money stock, it is the size of the money stock that determines the price level.
4. The money wage, *W*, adjusts to the equilibrium price level to produce the real wage required for full employment equilibrium. With the equilibrium real wage determined in Part B as $(W/P)_0$ and the equilibrium price level determined in Part D as P_0, the required money wage is determined in Part C as W_0.
5. As shown in Part E, saving, *S*, is a direct function of the interest rate; and investment, *I*, is an inverse function of the interest rate. With the interest rate as a measure of the reward for saving, the higher it is, the greater will be the volume of saving. With the interest rate as the "price" of capital goods, the lower the interest rate, the greater the volume of investment. The interest rate is determined by the intersection of the saving and investment functions, and it determines how real income is allocated between saving and consumption and how production (equal to real income) is allocated between consumption goods and capital goods.

These propositions are the basis for answers to the questions originally posed. They not only specify the levels of employment, output, consumption, saving, interest rate, price level, and money wage rate, but something much more: they show that the employment level is uniquely established at the full employment position. With both wages and prices flexible, which simply means that the money wage will fall if unemployment appears and the price level will

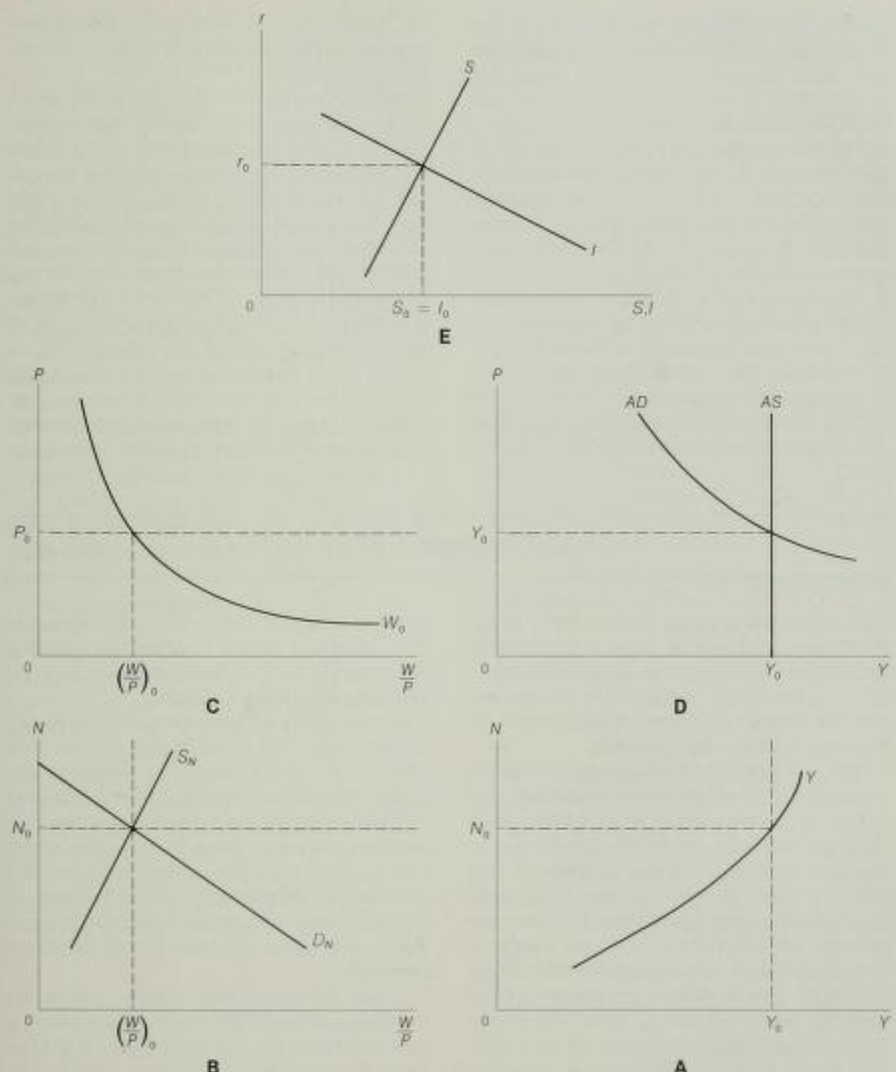


FIGURE 14-6
The classical model: including saving and investment

fall if the existing level of output cannot be sold at going prices, the automatic full-employment conclusion follows logically. It also follows that output will be that which can be produced with a fully employed labor force, that the price level will be that at which the money supply with its given velocity will purchase this full-employment level of output, and that the money wage will be so related to the price level as to make it profitable for employers to produce the full-employment level of output.

The classical conclusion that the economy has an automatic tendency to move toward a full-employment equilibrium is not widely accepted today. But this and other conclusions of classical employment theory can be rejected only by rejecting the assumptions on which that theory rests, for the theory itself appears to be internally consistent. Once its assumptions are granted, the theory inevitably leads to the indicated conclusions.

We will not enter into Keynes's specific attack on the assumptions that underlie classical theory, but most economists agree that his attack was successful. Not only did he offer persuasive arguments against these critical assumptions, but he replaced the rejected assumptions with others that appeared much more consistent with the facts of ordinary observation and statistical evidence. To the extent that the assumptions on which the classical theory of employment and output was based could be shown to be unacceptable, the conclusions, including the automatic full-employment conclusion, reached by that theory also became unacceptable.

A CONCLUDING NOTE

If the classical analysis of the process by which the levels of employment, output, and prices are determined is unacceptable, at least in its application to the modern economy, it may appear that this chapter is basically unnecessary. To this there are a number of replies.

First, it is not altogether correct to label the classical theory of employment, output, and prices as unacceptable or in some sense "wrong." Since the aim of this chapter was to merely introduce the broad outlines of that theory, it could do no more than draw broad conclusions and compare them with the conclusions so far derived from our study of Keynesian theory. The omission of refinements that would give us a more accurate picture of classical theory leaves us with little choice but to categorize the basic propositions of classical theory as correct or incorrect, and such categorization is itself inherently incorrect. Alfred Marshall once said that every *short* statement about economics is misleading (with the possible exception of this one). Our statement here, relative to what is involved in a complete treatment, is such a short statement and unavoidably somewhat misleading.

Second, one's understanding of a new theory is surely enriched when that theory is contrasted with the old theory that it seeks to displace. The classical system was the accepted explanation of macroeconomic phenomena for well over a hundred years. A discussion of this theory, which is partially correct and a product of the not so distant past, helps us understand and appreciate the changes in macroeconomic theory that have occurred since the Great Depression.

Finally, it is important to note that, despite the dramatic success of Keynesian theory over the past three decades, classical theory is still the theory in which many people in positions of great responsibility, both in government and in business, believe. It is not even necessary for them to have received formal training in eco-

nomics; the stuff of which economics is made has a way of permeating people's minds and influencing their outlooks without any awareness on their part. At the very end of the *General Theory*, Keynes expressed this thought in what has come to be a much-quoted statement:

> . . . the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.[13]

The defunct economists who continue to influence many of these people today are the economists who constructed the classical theory.

In a similar vein, we find the very last sentence in Alexander Gray's classic little handbook on economic thought:

> No point of view, once expressed, ever seems wholly to die; and in periods of transition like the present, our ears are full of the whisperings of dead men.[14]

For more than one reason, the teachings of the classical economists are of something more than historical interest forty-some years after the publication of the general theory. The few reasons here given should be sufficient to make this point. The "Keynesian Revolution" did not so completely wipe out the "old order" that no sign of it remains. Today's monetarism with the modern quantity theory at its foundation evolved from that "old order." Although it has been influenced by and incorporates some of the insights of Keynesian economics, in general outline it bears a close resemblance to that "old order." For both academic and practical reasons, a proper introduction to macroeconomic theory must still include the fundamentals of the theory that held sway for more than a century before Keynes.

[13] Keynes, *General Theory*, p. 383.

[14] A. Gray, *The Development of Economic Doctrine*, Longmans, Green, 1931, p. 370.

chapter 15

Money and the Rate of Interest

Although Keynes's classic carries the title *The General Theory of Employment, Interest, and Money*, it is possible to work part way through that theory, as we did in Part 2, without encountering money in a determinative role. Up to this point, money appears only as a common denominator or a convenient measure in which to state values and express flows. But at some point it must be recognized that money is a good deal more than this. This point is reached about halfway through the *General Theory*, where the following statement is found: "We have now introduced money into our causal nexus for the first time, and we are able to catch a first glimpse of the way in which changes in the quantity of money work their way into the economic system."[1] The "causal nexus" Keynes refers to is that between the supply of money and the rate of interest, the rate of interest, as we saw earlier, being a factor that influences the rate of investment spending and so influences aggregate demand.

The primary purpose of this chapter is to examine how the supply of and demand for money enter into the determination of the rate of interest. The theory to be discussed is a simplified version of the Keynesian liquidity-preference theory, which, as we shall see, differs sharply from the classical theory of the rate of interest discussed in Chapter 14. Although both the rate of interest and the supply of money are, of course, essential elements of classical theory, they are not tied together there as in Keynesian theory. The classical theory starts off with the proposition that the level of output will be whatever a fully employed labor force is capable of producing and supports this proposition as following naturally from the working of a system of flexible wages and prices. The rate of interest in classical theory was seen to depend basically on the "real" factors of the supply of saving ("thrift") and the demand for investment ("productivity of capital"). It determined how much of the full employment level of output would take the form of capital goods and how much the form of consumer goods, but the total amount of that output was quite independent of the interest rate. Finally, the money supply entered the classical

[1] John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936, p. 173.

system to determine aggregate demand but with output already established at the full employment level, aggregate demand merely establishes the price level at which that output sells in the market.

In the simple Keynesian theory, however, the level of output is made to depend on aggregate demand. Only if aggregate demand is sufficient will output be at the full employment level. If the level of output is below the full employment level, changes in the money supply will affect that level, because such changes affect interest rates and changes in interest rates in turn affect aggregate demand, primarily by influencing investment spending. In the Keynesian theory, the theory of money and interest thus becomes inseparable from the theory of output and employment. In this chapter we will concentrate on the theory of money and interest. Then, in Chapter 16, we will proceed to the theory that combines output and employment with money and interest.

MONEY AND OTHER ASSETS

According to the narrow definition of money, the only assets that qualify as money are currency and demand deposits. Each person on any date has some amount of wealth, and each person may hold wealth in a number of forms, including money, as here defined, interest-bearing securities such as bonds, equity securities or stock shares, real estate and physical assets of other kinds. He or she may distribute total wealth in various ways among these assets and will seek that particular portfolio which maximizes utility. The amount of any one asset in the portfolio may be expected to vary directly with the rate of return on that asset and inversely with the rates of return on substitute assets. For example, if the implicit rate of return on money stays the same while the rates on one or more alternative assets rise, portfolio balance calls for a shift toward a smaller proportion of total assets in the form of money and a correspondingly larger proportion in one or more other assets. Furthermore, given the large number of different kinds of assets available to be held and actually held by some persons, a change in the yield on any one of these will often involve more than an offsetting switch between this particular asset and another. To achieve the portfolio that maximizes the wealth-holder's utility may involve numerous changes.[2]

The approach to money and interest that takes into account all of the assets that a person may choose to hold as an alternative to holding money involves a major extension of the approach taken by Keynes. In his approach, which is the one we will be working through in this chapter, only one of the alternative assets is specifically taken into account: interest-bearing securities. There is, in other words, a focus on two assets: money and interest-bearing securities. A wealth-holder may at any time supply interest-bearing securities in the market and demand money, or he may supply money and demand interest-bearing securities. Although money is demanded for other reasons, the demand for money that is the opposite side of the supply of interest-bearing securities is critical in the theory of interest. Once we have worked through the concept of the demand for money, we will have an explanation of why the interest rate is what it is at any time with the money supply as given at that time.

[2] An introduction to the process of portfolio balance is provided in Chapter 25 as part of the analysis of how monetary policy works. See pp. 506–509.

THE DEMAND FOR MONEY

One reason people demand money is because money is needed in any economy in which almost everyone, persons as well as firms, sells goods and services (including factor services) in the market for money and uses money in turn to buy the goods and services offered by others. Functionally this amounts to the use of money as a medium of exchange. Classical theory explained the demand for money as essentially a demand resulting from this need for money as a medium of exchange.

In Keynesian theory, money becomes much more than a medium of exchange, much more than a device for mediating transactions in the marketplace. People also demand money for speculative purposes and as security against unforeseen needs for cash reserves. The breakdown of the demand for money into transactions and precautionary and speculative demands plays a vital part in the theory advanced by Keynes to explain the interest rate.

Transactions Demand

Everyone needs to hold some amount of money to carry out ordinary day-to-day transactions. However, the closer the synchronization between the timing of one's receipts and the timing of payments, the smaller the average money balance one will have to hold for this purpose. In the limiting case of a perfect coincidence between the amount that a person receives at each point in time and the amount that he or she pays out at each point in time, no money balance at all would be required for transactions. In practice, of course, no person or firm even approaches this limiting case, despite the ability of each to exercise some control over the timing of his or its receipts and payments. Everyone must, therefore, hold some amount of money to cover the unevenness between the timing of what comes in and what goes out.

No two persons have identical time patterns of payments, and few find their total payments for any month evenly distributed over the month. Particular days include above-average payments: the day the mortgage payment or the rent is due, the day the car payment must be made, the day for payment of all or part of the balance on department store charge accounts, Master Charge, BankAmericard, and so forth. However, as a first approximation, let us assume in what follows that there is an even distribution of payments over the month. What we want to see is that, with this pattern of payments or one approximating it, a person whose receipts for the month all come in on the first day of the month will require a larger average money balance over the month than would be the case if these same total receipts came in at intervals during the month. As a general rule, the average money balance a person or firm must hold over time for transactions declines as the frequency of his receipts rise.

As an illustration, assume an individual has receipts of $1,400 per month and that he makes payments of an equal amount each month. For convenience, assume that each month has exactly four weeks. Monthly receipts of $1,400 may come in in a number of different time patterns, such as $1,400 on the first day of the month, $700 on the first and fifteenth days of the month, or $350 on the first day of each week. (An amount of $2,800 bimonthly would also technically qualify.) If this individual actually receives the $1,400 on the first day of the month and disburses the full amount evenly over the month, he will be holding a money balance of $1,400 at the beginning of the first week, $1,050 at the beginning of the second week, $700 at the beginning of the third week, and $350 at the beginning of the fourth week.

His balance shrinks to zero at the end of the fourth week and then jumps back again to $1,400 the next day. His *average* money balance for the month in this case is $700, or an amount equal to half of his *monthly* receipts. His actual money balance at each point during the month is shown by the curve in Part A of Figure 15-1.

Suppose now that this individual receives the very same monthly receipts at a rate of $350 per week paid to him on the first day of each week. Assuming as before that his payments equal his total receipts, his money balance at the beginning of the first day of the week will be $350, at the beginning of the second day $300, and so forth with an *average* money balance of $175 for the week, an amount equal to one-half of his *weekly* receipts. In this case, his actual money balance at each point during each week is shown in Part B of Figure 15-1. The switch from an arrangement in which all of the month's income is received at the beginning of the month to that in which one-fourth of the same income is received at the beginning of each of the month's four weeks (4 × $350 instead of 1 × $1,400) means a fourfold rise in the frequency of receipts and a reduction in this person's average money balance to one-fourth ($175 instead of $700) of what it would be in the monthly case. If we extend the illustration to the case in which the monthly receipts of $1,400 come in at $50 per day, we would find that this sevenfold increase in the frequency of receipts (7 × $50 instead of 1 × $350) would be accompanied by a reduction in this person's average money balance to one-seventh ($25 instead of $175) of what it would be in the weekly case. In pushing this illustration all the way to the daily case, we reach a point of synchronization between each day's receipts of $50 and each day's payments of $50. The person holds an average balance of $25 during each day, but he does not carry any balance over from one day to the next as is required in the weekly case and *a fortiori* in the monthly case. The general rule noted earlier should now be clear: the average money balance a person or firm finds it necessary to hold in order to mediate transactions decreases as the frequency of his or its receipts increase.[3]

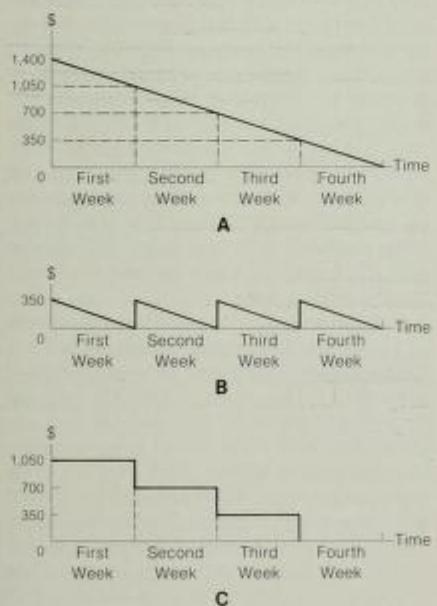


FIGURE 15-1
Hypothetical allocation of an individual's transactions balance between cash and earning assets

[3]It must be emphasized that this general rule applies on the assumption that the pattern of a person's *payments* over time is even or at least approximately so. A person's *purchases* may be spaced out evenly over time, but *payment* may occur at discrete points in time. Suppose a person's purchases of $50 per day are all on credit, with the settlement date of all accounts falling on a particular day each month. If his receipts of $1,400 per month all come in on the same day that payment of the $1,400 falls due, his average money balance for the month will be close to zero. He may hold money for only a matter of hours or minutes on that one day. On the other hand, if this person's receipts came in evenly at $50 per day, it will be seen that his average money balance for the month would be $700. His

This general rule holds true with the usual assumption that other things remain equal. In our illustration, a person's balance drops from \$700 to \$175 as he switches from monthly receipts to weekly receipts on the assumption that his total receipts and payments remain at \$1,400 for the period of a month. If these should somehow jump fivefold at the same time that there is a switch from monthly to weekly receipts, this person's average money balance would rise from \$700 to \$875. In this case, the effect that the rise in the dollar amount of receipts and payments has in increasing his average money balance more than offsets the effect that the switch from monthly to weekly receipts has in decreasing it. If this fivefold increase in receipts and payments had occurred with no change in the frequency of receipts and no change from the same even distribution of payments we are assuming, the person's average cash balance needed to handle the larger volume of transactions would be five times what it was before. This leads us to another general rule: The average money balance a person or firm finds it necessary to hold in order to mediate transactions increases proportionally with the dollar volume of transactions. Here again we hold other things equal, in this case including the frequency of receipts of each person and firm.

To sum up, over time the amount of money needed to handle transactions will tend to shrink to the extent that some persons and firms achieve a closer degree of synchronization between receipts and payments; over time the amount of money needed will tend to increase to the extent that the dollar volume of transactions to be mediated increases. The dollar volume of transactions has, of course, doubled and redoubled over the long run. Whatever the strength of the forces working in the other direction may have been, the rising volume of transactions has many times outweighed it to give us an almost uninterrupted increase year after year in the amount of money balances that all persons and firms combined find it necessary to hold in order to mediate the total volume of transactions.

Transactions Demand as a Function of Income The actual growth in the total volume of transactions has been accompanied by a growth in the size of the economy's gross national income or gross national product. Because the dollar volume of transactions for each time period includes all kinds of transactions in intermediate product and in securities and existing real property, it far exceeds each period's gross national income or product. However, in the belief that the ratio of GNP to the dollar volume of all transactions is reasonably stable, we may say as a first approximation that the amount of money balances that the public as a whole wishes to hold for transactions purposes depends directly on the level of income. We can thereby relate money balances to the variable with which we have been working all through the preceding chapters.

In this relationship, the actual dollar amount of money the public seeks to hold to carry out the transactions associated with any given level of output, Y, will vary proportionally with the price level, P, at which that output sells. It obviously calls for the use of twice as much money to buy one unit of some good when the price is \$100 that it takes to buy one unit of that good when the price is only \$50. The same is true for the grand total of purchases designated by PY. Any change in PY will require the same change in transactions balances whether that change is entirely in P, entirely in Y, or in any combination of the two. For example, if the public needs \$1 to handle the transactions represented by \$4 of income, required transactions balances will be \$100 billion when income is at the \$400 billion

balance would rise by \$50 per day to \$1,400 at the end of 28 days, at which time he would pay the \$1,400 which then falls due. In this case, the more frequent his receipts, the larger his average money balance. Although this is plainly an extreme illustration, it does indicate that there can be special conditions under which the general rule as stated does not hold.

level, whether that level of PY is \$4 × 100 billion units or \$1 × 400 billion units or any other combination equal to \$400 billion.

This relationship may be expressed in equation form as $M_t = k(PY)$, where M_t is the amount of money demanded for transactions, and, assuming M_t and (PY) are linearly related, k is the fraction of money income over which the public wishes to hold command in the form of transactions balances. Following the illustration above in which k is assumed to be 1/4, M_t will be \$100 billion when PY is \$400 billion. In the same way, M_t would be \$125 billion if PY were \$500 billion.

The amount of money demanded for transactions may be expressed in *nominal* or in *real* terms. The equation above, $M_t = k(PY)$, which is in nominal terms may also be written as $M_t = P \cdot k(Y)$ on the assumption that a change in P will bring about a proportional change in the dollar amount of transactions balances demanded, an assumption customarily made. Then, to convert from a demand for nominal balances to a demand for real balances, one merely divides $M_t = P \cdot k(Y)$ through by P to get $M_t/P = k(Y)$, where M_t/P is the amount of real balances demanded for transactions. This amount will also be designated by m_t.

The way in which the amount of these real money balances demanded varies with the level of Y and the size of k is illustrated in Figure 15-2. The line labeled $k(Y)$ indicates the amount of real balances, m_t, for various levels of Y with k assumed to be 1/4. If Y were \$400 billion, m_t would be \$100 billion; if Y were \$500 billion, m_t would be \$125 billion. If k were 1/5, we would have the line $k'(Y)$ in this figure. For real incomes of \$400 and \$500 billion, real transaction balances demanded would then be only \$80 and \$100 billion, respectively.

The size of k depends on institutional and structural conditions within the economy such as the degree of synchronization between receipts and payments for each person and firm, a factor emphasized above; on the time required for payments originating in one location to become receipts in other locations, or the speed of movement of money; on the degree of integration of industry, integration converting what were money payments between firms into mere intrafirm bookkeeping entries; and on other such factors. Some of these factors do not change significantly over a period of a few years, but it appears of late that the speed of movement of money does. Recent years have witnessed revolutionary changes in the technology of moving money. Developments like the checkless transfer of wages and salaries by computer, the lock-box system of mail delivery of checks, and the use by the Federal Reserve System of a fleet of jets to transport checks between Federal Reserve Banks have contributed to the fact that each dollar of demand deposits, on the average, mediated about two and a half times the dollar amount of transactions in 1974 that it did in 1964. In terms of Figure 15-2, this suggests that the slope of the straight-line may show an appreciable decrease over a few years. However, in order to simplify the present analysis, we will assume that k is stable in the short run. In other words, we will assume that the chief determinant of changes in the amount of real balances held for transactions is changes in the level of real income. Graphically, changes in m_t are then primarily the result of movements along a line like $k(Y)$ rather than changes in k, i.e., in the slope of the line.

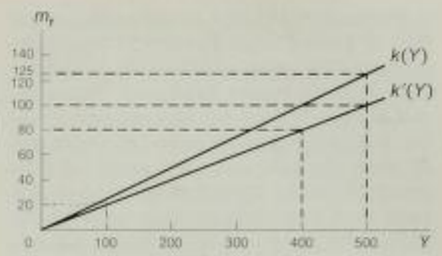


FIGURE 15-2
Transactions demand for money

Transactions Demand as a Function of the Rate of Interest It is plausible to expect the amount of real balances demanded for transactions purposes to vary directly with the level of real income. Perhaps it is also plausible to expect it to vary inversely with the rate of interest. Consider again the case of a person who receives $1,400 on the first of each month and spends his total receipts evenly over the 28 days of the month. As noted above, this person's holdings of money for transactions at any point during the month would be as shown in Part A of Figure 15-1. It may be seen that this person holds for transactions $1,050 of completely idle money during the first week, $700 during the second, and $350 during the third. Could he not convert this temporarily idle money into an interest-bearing asset? Parts B and C of the figure illustrate what he might do. On the first day of the first week (payday), $350 is retained to cover the first week's payments (see Part B), and $1,050 is used to acquire earning assets (see Part C). On the first day of the second week, he cashes in $350 of his earning assets to obtain money to cover payments of the second week, and so forth for the third and fourth weeks, after which the cycle repeats itself. In this fashion, the individual reduces his average transactions balances to $175 for the month (Part B) and makes his average earning-asset balance $525 for the month (Part C). These, of course, add up to the same $700 paragraph could clearly allocate his transactions balance in the way described by Figure 15-1. However, whether he would do so is another matter. To do so he would have to make four trips to the savings and loan association each month (and in addition four trips per month to a commercial bank if he uses a checking account to make payments). He might avoid these trips if deposit by mail and transfer to checking account by phone were available, but in any event he would have to weigh the interest earned against whatever time, effort, and inconvenience are involved. In these circumstances, some individuals would decide that it "pays" and others would decide that it does not.

For those who maintain a checking account, the time, effort, and inconvenience may be largely avoided at the cost of a somewhat lower rate of return by maintaining a savings account at the same bank. Many commercial banks permit an unlimited number of checks to be drawn with no service charge, if the customer maintains a minimum balance in a savings account. The individual in question can arrange for the automatic transfer of funds from his saving to his checking account to correspond with his disbursement of those funds; no minimum balance is required in his checking account. Although the rate of interest paid on his savings account with a commercial bank is lower than that paid on his account wi[illegible]

struments charge fees for their services in connection with each purchase and sale. The individual must here weigh the financial cost (and still some inconvenience) of frequent entry to and exit from the market for securities against the apparent advantage of holding interest-bearing securities in place of idle transactions balances. Among other things, the cost per purchase and sale, the rate of interest, and the frequency of purchases and sales determines the profitability of switching from idle transactions balances to earning assets.[4] Nonetheless, with the cost per purchase and sale given, there is clearly some rate of interest at which it becomes profitable to switch what otherwise would be idle transactions balances into interest-bearing securities, even if the period for which these funds may be spared from transactions needs is measured only in weeks. The higher the interest rate, the larger will be the fraction of any given amount of transactions balances that can profitably be diverted into securities.

During the 1973–74 period of extremely high short-term interest rates, many "money market" or "liquid asset" mutual funds sprang up. Through these funds, individuals without the means to purchase money-market instruments directly are able through a "pooling" of funds to participate indirectly in the market for these instruments with as little as \$1,000. However, financial costs are faced here also as the operators of such funds pass on their costs and add a charge for their services in the form of a "management fee." But as with the direct purchase of Treasury bills and the like, the higher the interest rate being paid on such instruments and therefore the higher the rate being paid by this kind of mutual fund, the more profitable it is for people to divert transactions balances into this kind of fund.

The relationship between the level of income, the rate of interest, and the transactions demand for money for the economy as a whole are depicted in Figure 15-3. If $Y = \$400$ billion and $k = 1/4$, m_t is \$100 billion as shown by the curve Y_1. This figure of \$100 billion, however, holds only as long as the interest rate is not above 4 percent, for example. As the rate rises above 4 percent, the transactions demand for money becomes interest elastic, indicating that, given the costs of switching into and out of securities, an interest rate above 4 percent is sufficiently high to attract some amount of transactions balances into securities. For still higher rates, the amount so diverted becomes larger, as indicated by the backward slope of the Y_1 curve. For a level of income of \$500 billion, the transactions demand curve shifts to Y_2 but again slopes backward at an interest rate above 4 percent. The curves Y_1 and Y_2 correspond to incomes of \$400 and \$500 billion on the $k(Y)$ curve of Figure 15-2.

It is difficult to generalize on the interest elasticity of the transactions demand for money for the economy as a whole. A giant corporation that this month holds millions of dollars not

[4] This point is examined in detail in J. Tobin, "The Interest-Elasticity of the Transactions Demand for Cash," in *Review of Economics and Statistics*, August 1956, pp. 241–47, and in W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," in *Quarterly Journal of Economics*, Nov. 1952, pp. 545–56. The following simple illustration will bring out the basic point. Assume an individual is holding \$1,000 for the purpose of paying a bill that falls due one month later. Even if he were able to buy so small an amount, would it pay for him to put this \$1,000 into a security for the one month? If the dealer imposes a flat charge of \$1 per purchase and sale plus an additional charge of 10¢ per \$100, it will cost our individual \$2, or \$1 + (10 × 10¢), to buy a \$1,000 security at the beginning of the month and an equal amount to sell at the end of the month. Total cost is \$4. If the interest rate that he can earn is 6 percent per annum, for one month he will receive 1/12 × 0.06 × \$1,000, or \$5, to leave him with a gain of \$1 over costs. At a 4-percent interest rate, he will earn \$3.33, or an amount less than his costs. A 4.8-percent rate is that at which the interest earned will just cover costs. It does not, however, follow that the individual will buy at any rate above 4.8 percent. With only \$1,000 involved, the \$1 to be earned at a 6-percent rate or even the \$2 to be earned at a 7-percent rate may not be worth the trouble of the record-keeping and paper work required. At the same time, if the amount involved is relatively large, a 6-percent rate will be quite attractive. On \$1 million for the month, there will be a gain of \$3,998. Even for the wealthy individual, this amount will offset the minor problem involved in arranging the purchase and sale of securities.

$m_t = k(Y)$

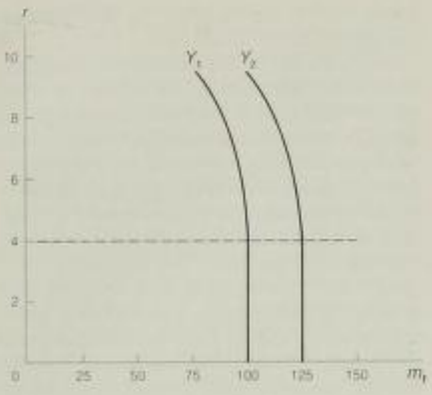


FIGURE 15-3
Interest-elastic transactions demand for money

needed for transactions until next month will not pass up the opportunity to put these funds into earning assets for a month or even less if the interest rate is high enough to permit a profit. An individual whose average transactions balance is moderate is less likely to be so interest-rate conscious. His transactions demand is apt to be completely interest inelastic at other than irresistibly high interest rates rarely found in actual experience.

Most economists agree that in practice there is some interest rate at which the transactions demand curve for money for the economy as a whole begins to slope backward (as in Figure 15-3). This means that our equation for transactions demand should become $m_t = f(Y, r)$ and that there is no longer a simple linear relationship between m_t and Y. To simplify our analysis, however, we will assume that this demand is perfectly inelastic with respect to the interest rate and retain our simple equation for this demand: $m_t = k(Y)$. In terms of Figure 15-3, the transactions demand for money becomes a function of the level of income alone. Changes in the level of income shift the demand curve, as from Y_1 to Y_2, but the curve is perfectly inelastic with respect to the interest rate at high as well as low rates.

Precautionary Demand

Transactions demand for money stems largely from a lack of synchronization between receipts and expenditures; similarly, precautionary demand arises primarily because of the uncertainty of future receipts and expenditures. Precautionary balances enable persons to meet unanticipated increases in expenditures or unanticipated delays in receipts.

This type of demand for money may be expected to vary to some extent with one's income. Individuals need more money and are better able to set aside more money for this purpose at higher income levels. The precautionary demand may also be expected to vary inversely with the interest rate. Unlike a transactions balance, which is something definitely scheduled for use in the near future, a precautionary balance is to secure one against a "rainy day" that may never come. At a high enough interest rate, one may be tempted to assume the greater risk of a smaller precautionary balance in exchange for the high interest rate that can be earned by converting part of this balance into interest-bearing assets.[5]

[5]Although precautionary demand may thus be expected to vary with the income level and the interest rate as transactions demand does, the individual's need that precautionary balances satisfy is one that may nowadays be met quite well at little or no cost in ways other than by holding money balances. For many individuals, the recent development of credit cards has reduced the amount of money they would ordinarily hold for precautionary purposes as well as for transactions purposes. Many emergency expenses can be met with a credit card—it can even be used to get cash immediately, and cash can, of course, be used to meet any kind of expenditure. For larger amounts, individuals can now do what business firms have long done: establish "lines of credit" with banks. Since a $1,000 line of credit means that this amount is immediately available at the individual's request, it is as if he had this amount sitting in his checking account as a precautionary balance. Although banks may charge a fee for setting up a line of credit whether the person draws on the line or not, some

Although precautionary demand may be formally distinguished from transactions demand, the total amount of money held to meet both demands is viewed primarily as a function of the level of income and, to some extent, of the interest rate. As in the case of the transactions demand, we will assume, for the sake of simplicity, that precautionary demand is interest inelastic and that it too depends solely on the level of income. If both transactions demand and precautionary demand are a function of income, the two may be combined so that our earlier equation, $m_t = k(Y)$, may now be understood to include in m_t both transactions and precautionary balances.

Speculative Demand

The proposition that money is held for transactions and precautionary purposes does not conflict with the classical view. A transactions balance is nothing more than money in its function as a medium of exchange, and a precautionary balance can be added to the classical system without materially affecting its conclusions. But this is as far as the classical theory went. The speculative demand for money, a systematic part of the demand for money in Keynesian theory, represents a distinct break with classical theory.

Classical theory assumed that a person would hold no money in excess of the amount needed to meet his transactions (including precautionary) requirements. To do so would be to forgo the interest that could be earned by putting that money into a security. The reasoning was that, even if the interest rate were very low, it would be better to get some return than none at all. Keynes pointed out, however, that one who buys a bond is "speculating" that the interest rate will not rise appreciably during the period in which he intends to hold the bond. If he believes that it will rise, he would be wise to hold non-interest-bearing money. It is this uncertainty as to the future interest rate that causes people to hold money for speculative purposes. If the future interest rate were known with certainty, there would be no speculative demand for money, and there could be no objection to the classical concept of the demand for money.

charge a fee only if one draws on the line. These banks, of course, limit this "free" service to those who maintain a deposit account with the bank; it is a device to attract new depositors to these banks, but to the individual it is a way of maintaining what amounts to a precautionary money balance at no extra cost to himself.

Security Prices and the Interest Rate To understand what is involved, we must examine the relationship between the interest rate and the market price of a debt security. Take, for example, a marketable U.S. government bond on which a purchaser faces no "credit risk"—that is, no risk of default or no risk that the interest and principal will not be paid as promised. There is no credit risk because the federal government can, if it so chooses, meet its obligations by merely creating the money needed for this purpose. Also for a U.S. government bond and for most other direct Treasury obligations, there is negligible "marketability risk" or "liquidity risk." An organized nationwide market exists for such securities in which a holder may at any time sell at a price close to the last quoted price. Unlike the markets for some other securities, the price variations on successive transactions are ordinarily minor. However, while a U.S. government bond is practically free of credit and marketability risks, it, like other securities, is subject to "market risk," the risk that the market interest rate may change.

These U.S. government bonds, like any other bonds freely traded in the market, have a dollar amount and an interest rate, commonly referred to as the coupon rate, printed on the face of the obligation. The number of dollars the issuer will pay each year to the holder is this coupon rate times the face amount. For example, a $1,000 bond with a 5-percent coupon promises to pay $50 interest each year over the life of the bond and the $1,000 face amount at matu-

rity. If we let $R_1, R_2, \ldots, R_n$ represent the number of dollars of interest to be paid on such a bond in years 1 through n, A represent the face or principal amount to be paid at maturity in year n, and r represent the rate of return currently being earned by security owners on bonds without credit risk and with the same maturity as the bond in question, then the current market value, V, of such a bond may be found from the equation:

$$V = \frac{R_1}{(1+r)} + \frac{R_2}{(1+r)^2} + \frac{R_3}{(1+r)^3} + \cdots + \frac{R_n}{(1+r)^n} + \frac{A}{(1+r)^n}$$

Notice that, apart from the last element, this is the same equation used to determine the present value of a capital good when $R_1, R_2, \ldots, R_n$ designated the stream of net income expected from that physical asset over its life, and r designated the rate at which this income stream was to be discounted.[6] Recall from that analysis that the higher the discount rate, the lower is the present value of that stream of income. The principle is the same here for a debt security. Instead of a stream of income produced by a capital good, we now have a stream of income produced by a bond. The rate at which this stream of interest income is to be discounted is the interest rate currently earned in the market on bonds of this type with this maturity. The higher this interest rate, the lower will be the market value of the bond, and vice versa. Thus, once given the stream of interest payments and the principal amount to be paid at maturity, the present value of this bond can change for essentially only one reason—a change in the interest rate. In contrast, corporate, state, and local bonds on which some credit risk exists can fluctuate in price as the market evaluation of credit risk varies. On such securities, there is, in other words, some risk that the R's and the A may not be paid as promised. There is also for many such securities a substantial marketability risk, the thinness of the market sometimes resulting in an inability to sell at a price very close to the price of the last transaction in the same security. However, if we limit ourselves here to bonds of the federal government, we find that changes in the interest rate are virtually the only cause of fluctuations in their prices.

The arithmetic relationship between the price of a bond and the interest rate is most clearly brought out by the consol, a type of security issued by the British government. This security promises only to pay a specified number of dollars in interest per year. It has no redemption value or maturity date; an investor can convert it into money only by selling it in the market to another investor. If market conditions are such that 5 percent is the rate currently being earned on securities of this type, a consol will sell for \$1,000 if it pays \$50 in interest each year. A buyer who pays \$1,000 will get a 5-percent return on his funds. In this special case of consols, the equation above reduces to:

$$V = \frac{R}{r}$$

or

$$\$1{,}000 = \frac{\$50}{0.05}$$

Yet, however attractive the 5-percent yield may appear, a prospective purchaser who believes that the interest rate will rise may be better off to hold his \$1,000 completely idle, as a speculative balance, rather than buy this security. Suppose, for example, that he believes the interest rate one year from now will be 5.26 percent. The security must then sell at the lower price that yields 5.26 percent to its purchaser, or $V = \$50/0.0526 = \$950$. To buy the security today and hold it for one year promises interest income of \$50 and capital loss of \$50 or (apart from tax considerations) neither a net gain nor a net loss. In other words, a prospective purchaser who holds his \$1,000 as idle

[6]See pp. 162–63. If a capital good had scrap value at the end of its life, this could be designated as A, and its present value would be computed as is the maturity value of a bond.

cash for a full year and then buys the security will be as well off as one who buys it today, if the rate does in fact rise to 5.26 percent. If a prospective purchaser anticipated that the rate a year hence would be anything above 5.26 percent, it would clearly be to his advantage to hold cash rather than buy the security; if the expected rate were anything less than 5.26 percent, there would clearly be a gain in buying the security rather than in holding the cash.[7] With 5 percent as the original rate, an expected rate of 5.26 percent is the critical rate that marks the difference between an expected net gain and an expected net loss. If the anticipated changes were larger, the purchaser could expect a clear-cut net gain or loss. Thus, if the expected rate were 6 percent, the price of the bond would fall to $833.33, with a net loss to the purchaser of $116.66 for the year. On the other hand, if the expected rate were 4 percent, the price of the bond would rise to $1,250, with a net gain of $300 for the year. Including interest and capital gain or loss, a 1-percent rise in the rate results in a negative yield of 11.66 percent, a 1-percent fall in a positive yield of 30 percent for the year.

The same inverse relationship between the interest rate and price applies to conventional debt securities with specified maturity dates. However, as compared with a consol, the closer a conventional security is to its specified maturity date, the less sensitive its price will be to changes in the interest rate. A rise in the rate from 5 to 6 percent reduced the market value of the consol, which pays $50 interest per year, from $1,000 to $833.33. For a $1,000 bond paying $50 interest per year and maturing in ten years, the same rise in the interest rate would reduce market value from $1,000 to $926.39. If it were five years from maturity, the value would fall from $1,000 to $957.88, and if it were only one year from maturity it would fall from $1,000 to $990.57.[8] These results could be derived by inserting the appropriate figures into the first equation above and solving, but they are found very easily in published interest tables.

In sum, combining interest rate, maturity, and market value, we see that the market value of a debt security is inversely related to the interest rate and that any given change in the interest rate will exert a greater effect on that market value the more distant the security is from maturity.

Expectations and the Interest Rate Anyone who buys a bond or other debt security (other than one of very short maturity) is unavoidably speculating to some extent on future changes in the interest rate and facing the possibility of a financial gain or loss that comes with such changes. Although other considerations enter, persons who switch at any time from money to bonds expect the interest rate to fall and bond prices to rise; they regard the present interest rate as "high" and present bond prices as "low." Those who switch from bonds to money hold opposite expectations; they regard the present interest rate as low and bond prices as high. Clearly, anyone who views the current interest rate as high or low must have some "normal" rate in mind against which the current rate is being compared. This concept of a normal rate is itself a changing thing, there being, for example, some tendency in this country to expect a higher normal rate under a Republican than under a Democratic administration. But in one way or another, wealth-holders develop a concept of what a normal interest rate is and

[7] See *General Theory*, p. 202.

[8] This assumes that 5 percent was the original yield on the consol and on the three securities with varying maturities. Since yields customarily vary directly with maturity, a yield of 5 percent on the consol might be accompanied by yields of, say, 4, 3, and 2 percent, respectively, on the 10, 5, and 1 year maturities. If the consol sold for $1,000 to provide a yield of 5 percent, the other three securities, each of which promises $50 in interest per year, would probably have sold at successively greater premiums over $1,000. But it still follows that the fall in value that would result from an equal rise in the interest rate on each maturity will be less for the shorter maturity than for the longer.

find the actual interest rate at times either high or low relative to that concept.

Given the notion of a normal rate, if wealth-holders view the current rate as high, they then expect a drop in the rate as it returns to normal. At this high rate, wealth-holders will accordingly spurn money and hold bonds. Not only do they thereby currently enjoy the high rate of return provided by bonds, but they can look forward to capital gains as bond prices rise with the anticipated fall of the interest rate to normal.[9] If, on the other hand, wealth-holders view the current rate as low, they anticipate a rise in the rate as it returns to normal. They accordingly spurn bonds and hold money. The penalty paid in interest forgone is relatively small when the interest rate is low; the prospective capital loss is relatively large if the rate should rise as expected from its low to its normal level. Holding idle money becomes the financially prudent policy.

We have seen that the amount of speculative balances that people will want to hold varies inversely with the interest rate. To accompany our equation $m_t = k(Y)$ for the transactions and precautionary demand for money, we must now write an equation for the speculative demand for money. If the amount of nominal balances held for speculative purposes is designated by M_{sp}, we have $M_{sp} = P \cdot h(r)$ for this equation. Unlike the relationship in the transactions-demand equation in which M_t varies directly with Y, the relationship between M_{sp} and r is inverse. As with the equation for transactions demand, the amount of speculative balances demanded may be expressed in real terms by dividing $M_{sp} = P \cdot h(r)$ by P. This yields $M_{sp}/P = h(r)$ or the amount of real speculative balances demanded is a function of the interest rate. For a shorter form, M_{sp}/P will be designated by m_{sp}.

In the case of the transactions demand for money, it seems reasonable to maintain that the nominal amount of such balances demanded will vary proportionally with the price level as shown in the equation $M_t = P \cdot k(Y)$. Other things being equal, the public simply needs proportionally larger (smaller) nominal transactions balances to handle the transactions associated with a given level of output when the price level of that output is higher (lower). However, should we expect the nominal amount of speculative balances demanded also to vary with the price level in the way given by $M_{sp} = P \cdot h(r)$? The answer here is not so clear-cut. To see what is involved, consider the case of a one-time, overnight doubling of the price level. If a wealth-holder's income and wealth also double in money terms, nothing has changed in real terms. One would expect that the wealth-holder in question would now want to hold twice the number of dollars in his speculative balance that he previously held at any given interest rate or that he would seek to adjust the nominal amount of that balance in line with the change in the price level so as to leave the real balance unchanged.

While this seems reasonable enough, one must also recognize that not all wealth-holders will find that their income and wealth increase in nominal terms proportionally with the general price level. At the extreme, suppose one's wealth were held entirely in currency and consols and one's only income was the interest

[9] It is, of course, possible for wealth-holders to view the current rate as "high" in the sense that it is somewhere above "normal" and at the same time to expect it to go still higher. Thus, before they will spurn cash and hold bonds, they must not only view rates as above normal but also expect that the next movement will be downward toward the normal level. Although wealth-holders may under certain circumstances expect one rise in rates to be followed by another, they will sooner or later expect this movement to reverse itself. The historical record does not show interest rates moving uninterruptedly to ever higher levels, and wealth-holders know the record. Note, however, that if they did indeed expect interest rate movements to be cumulative instead of self-reversing, each rise in the rate would cause a further shift out of bonds into money. This would give us a direct relationship between the amount of speculative balances held and the interest rate, the reverse of the relationship which will be found in the demand curve for speculative balances shown below in Figure 15-4. While expectations of a cumulative movement in interest rates may be held for a short period of time, at some point in time the expectations of a reversal will come to prevail.

earned on holdings of consols. An overnight doubling of the price level would leave one's nominal income unchanged but cut real income in half, and it would similarly leave the nominal value of one's wealth unchanged but reduce it in real terms by approximately the same degree that it reduced income. Whatever the speculative demand for money by a wealth-holder in these circumstances, it is not reasonable to expect that one will adjust the size of one's speculative balance proportionally with the change in the price level.

Between the extremes are, of course, intermediate cases, so that the answer to the question seems to be that the amount of speculative balances that wealth-holders as a group will want to hold at any particular interest rate will depend to a degree on the price level but that the amount may not vary proportionally with the price level. However, the usual treatment today is to show speculative balances varying in the way indicated by the equation above, $M_{sp} = P \cdot h(r)$, and that is the treatment adopted here.

Working with this equation in the form that shows the amount of real speculative balances demanded, $M_{sp}/P = m_{sp} = h(r)$, we may show the relationship graphically by the curve in Figure 15-4. The higher the interest rate, the smaller the amount of real balances that wealth-holders choose to maintain for speculative purposes. At some high interest rate, the curve indicates that they will hold no money in speculative balances. As drawn here, at a rate of 10 percent or higher the speculative demand curve in effect coincides with the vertical axis—m_{sp} is zero. This amounts to saying that all wealth-holders believe that the rate is so high that it can only fall or that even if it should creep still higher, the current high interest rate will more than offset any capital loss that may result. At this rate, no one prefers money to bonds; bonds become perfectly "safe." At the other end of the curve, speculative demand becomes perfectly elastic. In the present figure, this occurs at a rate of 2 percent, a rate so low that wealth-

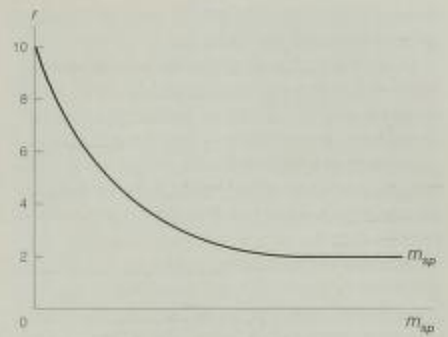


FIGURE 15-4
Speculative demand for money

holders believe that it can go no lower. To hold bonds at this interest rate instead of money is to face almost certain capital loss as interest rates rise and to find little offset against this loss in the interest income that is provided when the interest rate is so low. This section of the curve, known as the "liquidity trap," will be discussed later in the chapter.[10]

The focus here has been on Keynes's theory of the speculative demand for money, a theory we will be working with in what lies ahead but one with a serious shortcoming. In that theory, Keynes gives the wealth-holder a choice between holding risky bonds and riskless money. However, the wealth-holder in practice has the option of holding other debt forms that do not require that he incur risk in order to avoid the zero return on money holdings. He may, for example, hold wealth in time deposits, savings-

[10]Although we have here limited ourselves to Keynes's approach which stresses the concept of a normal rate of interest in explaining a downward-sloping speculative demand curve, later writing has shown that the demand for money as an asset will be interest elastic without resorting to the concept of a normal interest rate to arrive at this result. (See J. Tobin, "Liquidity Preference as Behavior Towards Risk," in *Review of Economic Studies*, Feb. 1958, pp. 65–86, reprinted in M. G. Mueller, ed., *Readings in Macroeconomics*, 2nd ed., Holt, Rinehart & Winston, 1971, pp. 173–91.)

and-loan shares, commercial paper, Treasury bills, and similar forms, all of which provide a rate of return, all of which in practice are immediately convertible into money, and all of which are either absolutely fixed in dollar value, regardless of changes in the market interest rate, or, in the case of a debt form like Treasury bills, are of such short term as to be virtually unaffected in value by changes in the market ~~interest rate~~.

~~The fact that~~ such riskless, short-term, interest-bearing debts are available to wealth-holders makes it difficult to believe that wealth-holders have a speculative demand for money in the way described by Keynes. For in Keynes's system this is money held specifically because the alternative asset (bonds) carries the risk of

Consider the opposite situation, in which the interest rate on long-term debt rises to a level that wealth-holders regard as above normal. They then anticipate a fall in the long-term interest rate, which will give them a capital gain on any holdings of long-term debt, something they will not get on holdings of short-term debt. For this reason, they tend to switch out of holdings of short-term debt into long-term debt. This raises the price of long-term debt and reduces the yield on such debt. Therefore, in this case, one finds a tendency for short-term yields to rise above long-term yields.[11]

If a movement of the long-term interest rate to a level above or below what wealth-holders regard as normal caused wealth-holders to shift out of or into short-term debt rathe

right. In what follows, we could, therefore, discard completely the Keynesian speculative demand for money and work only with an interest-elastic transactions demand for money, and we would still reach results similar to those reached with a speculative demand for money included. However, in order to trace the theory as it was developed by Keynes, we will follow the conventional procedure, which treats the transactions demand as interest inelastic and which includes an interest-elastic speculative demand for money. The transactions demand is thus $m_t = k(Y)$, and the speculative demand is $m_{sp} = h(r)$.

Total Demand for Money

The total demand for money expressed in real terms may be designated by m_d, where $m_d = m_t + m_{sp}$. Therefore, combining the equation $m_t = k(Y)$, which we understand to include precautionary demand, and the equation $m_{sp} = h(r)$ we now have an equation for the total demand for money:

$$m_d = k(Y) + h(r)$$

For any given price level, we know from k what m_t will be for each level of Y. Similarly, for any given price level, we know from h what m_{sp} will be for each level of r. We therefore know from k and h what the total demand for money will be for every possible combination of Y and r. This may be shown as in Figure 15-5.

Part A of the figure shows the transactions demand for money as \$100 billion when the level of income is \$400 billion, assuming that k is 1/4. Part B shows the speculative demand for money as an inverse function of the interest rate. Part C shows the total demand curve for money, the sum of the separate demands of Parts A and B, or the sum of m_t and m_{sp}. For example, at an income level of \$400 billion and an interest rate of 4 percent, total money demanded is \$110 billion; at the same income level but with an interest rate of 6 percent, total money demanded is \$105 billion.

The Supply of and the Demand for Money

Regardless of the demand for money, the nominal amount of money that people and firms hold at any time clearly cannot exceed the stock of money in the system at that time. Furthermore, it cannot be less. Whatever the stock of money may be at any time, someone must hold that amount.

Equilibrium in the market for money requires that the supply of money equal the demand for

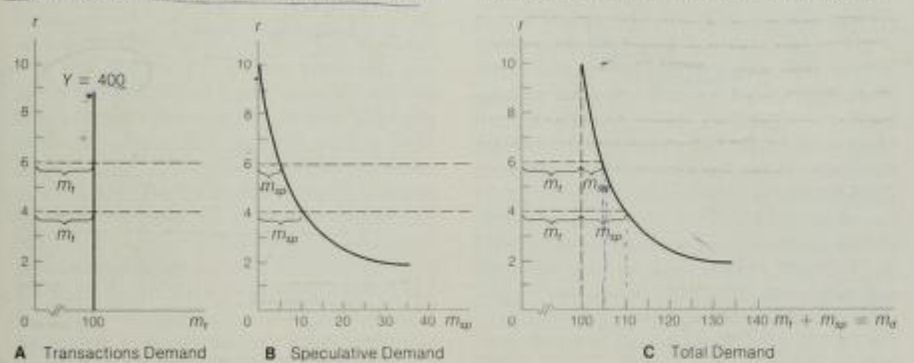


FIGURE 15-5
Total demand for money

money. Thus, if M_s represents the nominal amount of money that exists, equilibrium requires that $M_s = M_d$ and, in real terms, that $m_s = m_d$.

The nominal money stock, of course, changes over time in accordance with the policy decisions of the central bank that controls it, and the real money stock corresponding to any nominal money stock changes with the level of prices. Later in this chapter we will discuss briefly how and why the central bank may choose to increase or decrease the money supply. For the moment, let us assume that the real stock of money is some given amount. This then appears in Figure 15-6 as the perfectly inelastic supply curve, m_s. The demand curve, $m_d = (m_t + m_{sp})$, is carried over, with the scale expanded but otherwise unchanged, from Part C of Figure 15-5.

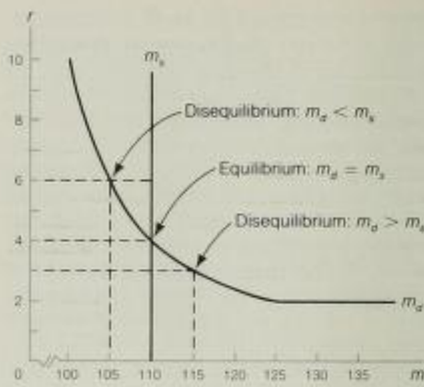


FIGURE 15-6
Equilibrium between supply of and demand for money

THE EQUILIBRIUM INTEREST RATE

Given the money supply and the income level, there is some particular interest rate at which the sum of the transactions and speculative demands for money will just equal the supply of money. The interest rate that equates the supply of and demand for money is the equilibrium interest rate. When the supply of money is \$110 billion, as in Figure 15-6, only at an interest rate of 4 percent is the amount of money demanded equal to the amount supplied. At a higher rate, say 6 percent, disequilibrium occurs, since money demanded is \$105 billion and money supplied is still \$110 billion. Because the total money supply must be held by someone, the public finds in such a situation that its actual holdings of money exceed the desired amount. After allowing for \$100 billion required for transactions, people find that the \$10 billion remaining is more than they choose to hold in the form of idle money balances at so high an interest rate.

People will therefore enter the market to buy securities with the excess cash. The increased demand for securities will drive the prices of securities up and reduce their yield. This will continue until security prices have been pushed up by the amount necessary to reduce their yield to 4 percent, at which point people will be content to hold the \$10 billion of speculative balances they actually do hold. Conversely, at any rate below the equilibrium rate, say at 3 percent, there is also a disequilibrium, but in this case one in which money demanded is \$115 billion and money supplied is only \$110 billion. At this low interest rate, people would rather hold less in securities and more in cash. Therefore, they try to sell securities and get into cash. The increase in the supply of securities drives down their prices and raises their yield. This continues until security prices have fallen by the amount necessary to raise their yield to 4 percent, at which point people will be content to hold the \$10 billion of speculative balances they actually do hold.

↑ MS → ↓ r → ↑ Inv. spending → ↑ Y → ↑ m_t → ↑ r

CHANGES IN THE EQUILIBRIUM INTEREST RATE

The interest rate, which is the price of money, will, like other prices, rise or fall in response to changes in supply and demand. Assuming no off-setting change in the price level, the real supply of money changes as government increases or decreases the nominal stock of money. On the same assumption, total demand changes as the level of income or the speculative demand for money changes. As a first step, we may therefore trace changes in the interest rate to changes in the money supply, the level of income, and the speculative demand for money. Changes in one of these variables may influence the value of another. For example, an increase in the money supply will reduce the interest rate, stimulate investment spending, and raise the level of income, thereby increasing the transactions demand for money and raising the interest rate above the level to which it fell as a result of the increase in the money supply. How all these interdependent variables fit together in a general model will be explained in Chapter 16. At this point, we will consider one change at a time, on the assumption that other things remain equal.

Changes in the Money Supply

The nominal money supply, M_s, was defined earlier as the total currency and demand deposits held by the public. Many factors affect the size of this total; one of the most important is the change in Reserve Bank holdings of government securities. The Reserve authorities can raise or lower the money supply through "open-market operations," the name for transactions that alter Reserve Bank holdings of government securities.[12]

Reread this section ?

The Process of Monetary Expansion and Contraction On any date, the balance sheets of the Federal Reserve Banks, the commercial banks, and all other firms and individuals collectively described as the public will reveal assets and liabilities measured in tens of billions of dollars. Our interest is only in the changes that occur in particular asset and liability items between two dates; such changes are indicated by + and − signs in the three balance sheets of Table 15-1.

Suppose that the Reserve authorities decide to increase the nominal money supply by \$10 billion. To do this the Reserve Banks purchase \$2 billion of government securities in the open market. The sellers of these securities are assumed to be individuals and firms (excluding commercial banks) that are part of the public. The Reserve Banks issue checks drawn against themselves in order to pay the sellers. The sellers of the securities deposit these checks at their commercial banks, thereby taking payment in the form of credit to their checking accounts; the commercial banks in turn send these checks to the Reserve Banks, taking payment in the form of a \$2 billion increase in their deposit balances with the Reserve Banks. These transactions are summarized as Step 1 in the table. At this step, the public has increased its money holdings by \$2 billion and decreased its security holdings by a like amount. The commercial banks, in turn, have increased their deposits with the Reserve Banks by \$2 billion and their deposit liabilities to the public by a like amount. The Reserve Banks have increased their security holdings by \$2 billion and, in payment, have created a like amount of deposit liabilities. These deposit lia-

[12] The factors determining changes in the money supply are, of course, examined in detail in texts on money and banking. See, for example, L. V. Chandler and S. M. Goldfeld, *The Economics of Money and Banking*, 7th edition, Harper & Row, 1977, Chs. 5–6 or E. Shapiro, *Understanding Money*, Harcourt Brace Jovanovich, 1975, Chs. 3–5.

bilities amount to an addition to the legal reserves of the commercial banks.

Suppose that commercial banks are legally required to maintain a minimum reserve of, say, 20 percent of their demand deposit liabilities in the form of deposits at the Reserve Banks. If, prior to Step 1 above, the commercial banks had just the amount of reserves required to back up their existing deposit liabilities, they would now have excess reserves of $1.6 billion as a result of Step 1. Such reserves are nonearning assets, and the commercial banks proceed to put them to work—by buying securities, for example. The sellers of securities—that is, the public—receive in payment checks drawn by the various commercial banks against themselves as buyers; these checks are deposited at various commercial banks. Step 2 in Table 15-1 shows that commercial banks as a group acquired $8 billion in securities (assets) and created $8 billion of new demand deposits (liabilities) in the process. Step 2 also shows an increase of $8 billion in the public's demand deposits, or money holdings, and a decrease of the same amount in its security holdings.

The net effects of the original open-market purchase of $2 billion in securities by the Reserve Banks are as shown in the last part of the table. If we assume that there is a 20-percent reserve requirement and that banks expand deposits to the limit set by their reserves, commercial bank reserves are up by $2 billion, which, together with the $8 billion in securities, supports the $10 billion of new demand deposits held by the public. In exchange for this $10 billion of new deposits, the public has given up a similar amount of security holdings, $2 billion of which is held by the Reserve Banks.

TABLE 15-1
An expansion of the money supply

	FEDERAL RESERVE BANKS		COMMERCIAL BANKS		PUBLIC	
	Assets	*Liabilities*	*Assets*	*Liabilities*	*Assets*	*Liabilities*
Step 1	Securities + $2 . . .	Deposits + $2 . . .	Reserves + $2 . . .	Public's Deposits + $2 . . .	Securities − $2 Deposits + 2	. . . . . .
Step 2	. . . . . .	. . . . . .	Securities + $8 . . .	Public's Deposits + $8 . . .	Securities − $8 Deposits + 8	. . . . . .
Net effects	Securities + $2 . . .	Deposits + $2 . . .	Reserves + $2 Securities + 8	Public's Deposits + $10 . . .	Securities − $10 Deposits + 10	. . . . . .

Demand deposits being money and securities not being money, there has been an increase of \$10 billion in the nominal money supply, or the banking system has "monetized" \$10 billion worth of securities.

If, instead of buying \$2 billion in securities, the Reserve Banks had sold to the public \$2 billion from its portfolio, the effects would have been exactly reversed. In Table 15-1, each + sign would become a − sign, and vice versa. With the public now purchasers instead of sellers of securities, there would be a \$10 billion decrease in the public's holdings of money and a \$10 billion increase in its holdings of securities.

The results described here assume that the amount of the deposit liabilities of the banks is independent of the interest rate that banks can earn on loans made or securities purchased. In other words, the banks are assumed to be fully "loaned up" at all times, so that any increase or decrease in their reserves will mean an expansion or contraction of their deposit liabilities by a fixed multiple of the change in reserves. Actually, banks do carry some excess reserves, and an aspect of this behavior that is relevant here is the fact that the amount of excess reserves they hold tends to vary inversely with the interest rate. This is similar to the behavior of individual wealth-holders with respect to the amount of speculative balances they choose to hold at different interest rates. We thus find that, with no change in the total amount of reserves held by the banks, the amount of their deposit liabilities outstanding will vary directly with the interest rate. Graphically, instead of a supply curve perfectly inelastic throughout as in Figure 15-6, the supply curve slopes upward to the right. If we allow for this aspect of bank behavior, we find that the supply of money as well as the demand for money depends on the interest rate. However, since the supply elasticity is not crucial in what follows, we will for simplicity assume as earlier that the supply of money consistent with any given amount of bank reserves is interest inelastic.[13]

Monetary Expansion and the Interest Rate The m_s and m_d curves of Figure 15-7 are the same as those of Figure 15-6. As we saw there, with demand as given, the money supply of \$110 billion produces an equilibrium interest rate of 4 percent. The m_{s_1} curve shows the supply of money after the \$10-billion increase that results from the expansion of bank credit. The increase in the supply of money forces the interest rate down to 2.5 percent. As long as there is no shift in the total demand curve for money, further increases in the supply of money will continue to lower the interest rate, and decreases will raise it. Although our model is highly simplified, it shows in general how the

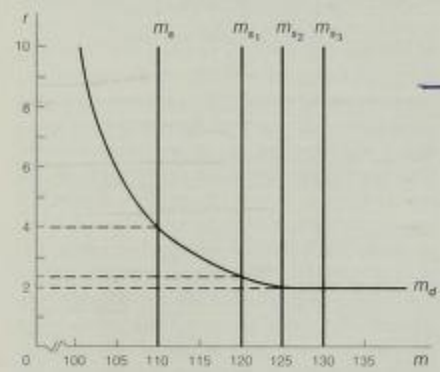


FIGURE 15-7
Changes in the supply of money and the rate of interest

[13]For an introduction to the relationship between the interest rate and the supply of money, see W. R. Hosek and F. Zahn, *Monetary Theory, Policy, and Financial Markets*, McGraw-Hill, 1977, pp. 90–103, and R. L. Teigen, "The Demand for and Supply of Money," in W. L. Smith and R. L. Teigen, editors., *Money, National Income and Stabilization Policy*, 3rd edition, Irwin, 1974, pp. 87–91.

Reserve authorities can raise and lower the interest rate by their ability to produce changes in the money supply, changes that are usually brought about by means of open-market operations as described above.

This ability rests in turn on the ability of the Reserve Banks to find sellers when it seeks to buy securities and to find buyers when it seeks to sell securities. As long as the Reserve Banks do not try to buy low and sell high in order to show a profit or avoid a loss in open-market transactions, there will be no shortage of buyers or sellers, whatever the scope of Federal Reserve operations in the open market. In Step 1 of our illustration, the Reserve Banks bought $2 billion in securities. Some rise in security prices may have been necessary to induce security-holders to swap interest-bearing securities for non–interest-bearing money. The rise in security prices would have been that needed to cause a sufficient number of security-holders to believe that security prices were high and likely to fall and that interest rates were low and likely to rise. When $8 billion more in securities were purchased by the commercial banks in Step 2, some further rise in security prices may have been necessary to induce the exchange of this amount of interest-bearing securities for money. In terms of Figure 15-7, for the complete sequence that involved an increase in the money supply of $10 billion, security prices had to rise enough to force the interest rate down from 4 to 2.5 percent. Only at this lower interest rate were people content to hold the additional $10 billion of money in place of an equal amount of securities.

The "Liquidity Trap" How much of a change in the interest rate may be expected from any specific increase in the money supply depends, other things being equal, on the elasticity of the speculative demand curve. If it is more elastic over the range from 4 to 2.5 percent than the m_a curve of Figure 15-7, the decrease in the interest rate effected by the $10-billion increase in the money supply would be less. If less elastic, it would be more.

There may be some relatively low rate (2 percent in Figure 15-7) at which the curve becomes perfectly elastic, indicating that the expectations of wealth-holders are virtually unanimous that the interest rate is so low that it can go no lower and that security prices are so high that they can go no higher. Although the Reserve authorities could, through open-market operations, expand the money supply past m_{s_2} to m_{s_3}, they could not in so doing succeed in reducing the interest rate below that set by the "liquidity trap." At this rate, the demand of wealth-holders for money is perfectly elastic, or the wealth-holders' supply of securities is perfectly elastic. The Reserve Banks could buy more and more securities in the open market, but the prices paid for them would stay the same. Monetary expansion is completely incapable of reducing the interest rate below the rate set by the liquidity trap.

Although it is an interesting and presumably possible phenomenon, the actual appearance of a liquidity trap is obviously a rarity. Seldom do interest rates reach the low level at which wealth-holders hold the expectations necessary to produce a liquidity trap. The closest approximation to a liquidity trap in U.S. experience was during the years immediately following the Great Depression, a period now over forty years behind us.[14] Although it is now viewed as an extreme case, the liquidity trap plays an important part in Keynes's *General Theory*, interestingly enough written at a time when the existence of a liquidity trap seemed more of a reality than a mere possibility. We will

[14]The existence or nonexistence of the Keynesian liquidity trap has been investigated by J. Tobin in his study, "Liquidity Preference and Monetary Policy," in *Review of Economics and Statistics*, May 1947, pp. 124–31, and in a more sophisticated form by M. Bronfenbrenner and T. Mayer, "Liquidity Functions in the American Economy," in *Econometrica*, Oct. 1960, pp. 810–34. See also D. Laidler, "The Rate of Interest and the Demand for Money—Some Empirical Evidence," in *Journal of Political Economy*, Dec. 1966, pp. 543–55.

have more to say about the liquidity trap when we develop an extended model of income determination in Chapter 16.

Changes in the Level of Income

If we take the money supply and speculative demand for money as given, the interest rate will vary directly with the level of income. Handling the larger dollar volume of transactions associated with a higher level of income calls for larger transactions balances. Let us retain the assumption that $k = 1/4$; with income at \$400 billion, the m_d curve of Figure 15-8 is the same as that in preceding figures. A rise in income from \$400 to \$420 billion raises transactions demand from \$100 to \$105 billion. As long as there is no change in speculative demand, the total demand curve shifts \$5 billion to the right at each interest rate to produce the new total demand curve, m_{d_1}. With an increase in the demand for money and no change in supply, the interest rate rises to a new equilibrium level of 6 percent.

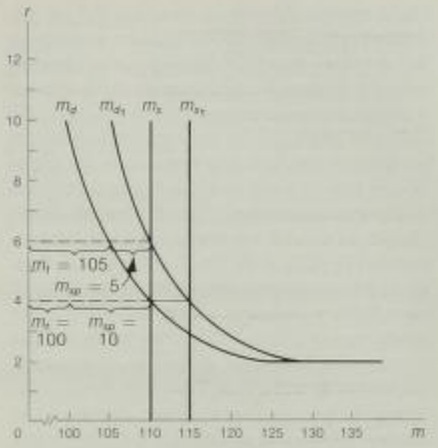


FIGURE 15-8
Changes in the income level and the rate of interest

The underlying process by which this rise in the interest rate takes place is not revealed by Figure 15-8. Basically, the process involves the diversion of money from speculative to transactions balances. As people and firms find that more money is needed to handle the greater volume of transactions accompanying a rise in income, they sell some of their security holdings in the market in order to secure the additional transactions balances needed. Since we are assuming that there is no change in the total supply of money, the additional transactions balances can come only from speculative balances. This transfer will occur as the prices of securities fall and their yields rise, the result of the increase in the supply of securities offered on the market. With the fall in prices and rise in yields, holders of speculative balances who were fearful of bonds at higher prices and lower yields will be tempted to switch to bonds at the more attractive price-yield combination now available. In our illustration, security prices must fall far enough to raise the interest rate from 4 to 6 percent in order to draw the required \$5 billion out of speculative balances into transactions balances.

By the same reasoning, with the money supply given, a fall in the income level means a decline in the interest rate. When the income level falls, the public discovers that it is holding more in transactions balances than it needs. This excess money will seek a return in securities, thus pushing up the prices of securities and reducing their yields to the point at which the public willingly holds the larger, idle money balances. The results of a decline in income from \$420 to \$400 billion are indicated in Figure 15-8 by a shift in the total demand curve from m_{d_1} to m_d. The equilibrium interest rate falls from 6 to 4 percent.

We are now in a position to tie together changes in the money supply and changes in the income level. As a rise in real income and

the accompanying rise in employment are desirable, public policy in this situation will oppose a rise in the interest rate because of the adverse effect on investment spending and on the further expansion of real income that it will have. In the example given above of a \$20 billion rise in income, open market purchases of \$1 billion of securities by the Reserve Banks would (on our earlier assumptions) provide the additional commercial bank reserves necessary for an overall increase of \$5 billion in the money supply. This increase would provide the additional transactions balances called for by the \$20 billion rise in income. There would be no rise in the interest rate and no danger of restraining the expansion of income through a restrictive monetary policy. This increase in the money supply is depicted in Figure 15-8 by the shift in the supply curve from m_s to m_{s1}, and as a result the interest rate is held constant at 4 percent.

In the opposite situation, a decline in income, public policy will favor a fall in the interest rate as a means of stimulating investment and arresting and possibly reversing the decline in output and the decline in employment that accompanies it. The model shows that some "easing" of the interest rate occurs without any action at all by the Reserve authorities. The decline in income will of itself automatically bring some reduction in the interest rate, but this will only prevent the decline in income from being greater than it otherwise would have been. For the typical downturn the Reserve authorities will go beyond a passive role and pursue an active, expansionary policy: more money will be pumped into the system as a step to bring about an upturn.

Changes in the Speculative Demand for Money

In our earlier discussion of Keynes's speculative demand, we noted that wealth-holders develop a concept of what are "normal," "low," and "high" interest rates. This is not intended to mean that all wealth-holders have the same opinion as to what high, low, and normal rates may be at any particular time. There will be and must be diversity of judgment if the concept of a speculative demand function is to have any meaning.[15] Despite this diversity of judgment among individual wealth-holders, there will also be a consensus among wealth-holders as a group. References to high, low, and normal rates are to what average opinion holds these rates to be at any given time.

If for some reason the consensus is that the normal rate will rise from the neighborhood of 4 percent to the neighborhood of 6 percent, there will be a movement out of securities into money, which will raise the interest rate to 6 percent. If there is no decrease in the money supply and no rise in the level of income, the interest rate will in fact rise if wealth-holders expect it to and if they act on their expectations. The same is true in the other direction. If there is no increase in the money supply and no fall in the level of income, the interest rate will fall from, say, the neighborhood of 4 percent to that of 3 percent, if wealth-holders expect it to fall and act on their expectations.

In Figure 15-8, with money supply given by m_s and demand for money given by m_d, the actual rate is 4 percent; we may assume this

[15] If all wealth-holders held identical views as to the normal interest rate at any particular time, and if each went "whole hog" into bonds at any rate above this rate and "whole hog" into cash at any rate below it, the speculative demand curve would be a perfectly elastic function at the level of this normal rate. Regardless of the size of the money supply, as long as it was greater than that needed for transactions purposes, the interest rate would remain unchanged. With the actual rate equal to what wealth-holders regard as the normal rate, there would be no reason for this perfectly elastic function to shift upward or downward and therefore no way for the interest rate to change in response to any change in the money supply. Since the interest rate does actually change, it follows that the speculative demand curve must be less than perfectly elastic if it is to have any meaning. This does not conflict with the possibility that a portion of the curve may be perfectly elastic. A liquidity trap may exist, but the whole curve cannot amount to a continuous liquidity trap.

is regarded by wealth-holders as in the neighborhood of the normal rate. If now for some reason general opinion should come to hold that rates near 6 percent will be normal, rates above the neighborhood of 6 percent will be viewed as high and rates below the neighborhood of 6 percent will be viewed as low. This amounts to an upward shift in the whole demand curve from m_d to m_{d_1}. With no change in the money supply or in the level of income, a new equilibrium interest rate is established at 6 percent.

It is easy enough to trace the simple mechanics of the process by which a change in the equilibrium interest rate will follow from a change in wealth-holders' opinions as to what the normal interest rate will be. It is not at all easy to explain specifically what causes wealth-holders' opinions to change. To begin with, it seems clear that the particular opinions concerning normal interest rates at any time cannot be separated from the actual rates observed over the near and more distant past. In an economy that has never experienced a rate on high-grade long-term bonds above 8 percent, for example, wealth-holders are not likely to come to believe that a rate above 8 percent can be a normal rate. The same is true on the low side. But within a range so defined there still appears to be room for considerable variation in opinion as to what the normal rate will be from time to time. Though it is only part of an answer, one important determinant is wealth-holders' views concerning future Federal Reserve policy. If there were no offsetting Federal Reserve action, one would usually expect the interest rate to begin to rise at some point in a period of vigorous business expansion following a recession low. Unless met by an expansion in the money supply, the increased transactions requirements would tend to force the interest rate up, as in U.S. business cycles before World War II. But compare this, for example, with the experience during the first four years of the economic expansion that began early in 1961. GNP at annual rates in current dollars rose by almost $160 billion from the first quarter of 1961 to the first quarter of 1965. However, the interest rate on long-term U.S. bonds fluctuated over a range of less than 0.5 percentage point during this period. As a result of Federal Reserve action to prevent the long-term rate from rising sharply, this rate, which averaged 3.8 percent during the first quarter of 1961, averaged no higher than 4.1 percent during the first quarter of 1965, despite four continuous years of business expansion. What does experience like this mean to wealth-holders' views of the normal interest rate? Changes in such views now become dependent on (1) the expected changes in all factors, domestic and international, that enter into the forecast of the coming business situation and (2) the expected response of the Reserve authorities to such changes, if changes actually come to pass as expected. Expected changes that would otherwise lead wealth-holders to anticipate a higher normal rate will not cause an upward shift in the speculative demand function if wealth-holders also expect that Federal Reserve policy will offset the rise in the rate that would otherwise occur.

An interesting result of this is the way the Reserve authorities may exploit wealth-holders' recognition of their power to maintain the interest rate unchanged or to raise or lower it over limited periods of time. If, for example, a rise in the rate is deemed necessary by the Reserve authorities, instead of actually selling securities in the open market, they may achieve the desired rise through public statements by the Chairman of the Board of Governors and other high officials of the Federal Reserve, warning that inflation once again represents a dangerous threat to the stability of the economy. Wealth-holders will interpret this as tantamount to an announcement of a pending "tight money" policy. The higher rate that the Reserve authorities seek may then be produced simply as a result of the upward shift in the speculative demand for money that will occur as wealth-holders act on their understanding of the real meaning of

the statements made by the Reserve authorities. In this way, the Reserve authorities may accomplish, with nothing more than a few words, what could otherwise be accomplished only by positive open-market action.

To illustrate in terms of Figure 15-8, let us assume that the Reserve authorities have planned to bring about a desired tightening with the money supply unchanged at m_s. With the equilibrium rate at 4 percent, all that may be needed is to make wealth-holders believe that the Reserve authorities will raise the rate to 6 percent. This expectation can in itself produce a shift from m_d to m_{d_1} in the demand function, which will yield the desired rise in the rate from 4 to 6 percent. The Federal Reserve can thus achieve its goal without actually intervening in the market.

THE DEMAND FOR MONEY—FROM CLASSICAL TO POST-KEYNESIAN THEORY

The simple quantity theory of money carries with it a simple theory of the demand for money. People supply goods and services in the market and demand money rather than other goods in exchange, since only money has the unique characteristic of being generally acceptable in exchange for any and all other goods and services offered for sale in the market. As money is received, it is spent. The number of times per time period that the existing money supply appears in the market as demand for goods and services is its velocity, which, for reasons discussed earlier, the simple quantity theory regarded as stable in the short run. For a system whose output was at the full-employment level, output too was stable in the short run. This meant that the price level varied proportionally with the money supply. Any additional money flowing into the hands of the public through a money-creating expansion of bank credit would in turn flow into the market for goods and raise the price level of goods proportionally. Thus, it could be said that the price level equates the supply of and demand for money. Any rise in the supply of money will raise the price level by the amount necessary to create an increase in the demand for money equal to the increase in supply. The higher P is, the greater is PY; the greater PY is, the greater is the amount of M required to purchase PY with a stable V.

The concept of the demand for money does not emerge very clearly when the quantity theory is approached, as it was here, in terms of the $MV = PY$ equation. Although an equilibrium between M_d and M_s is implicit in the equation, it is only implicit. The concept of the demand for money is hidden from view. If we rewrite this equation to read $M = 1/V(PY)$ and then substitute k for $1/V$, we have $M = k(PY)$. $MV = PY$ is known as the velocity formulation and $M = k(PY)$ as the cash-balance formulation.[17] Since $V = 1/k$ and $k = 1/V$, the two formulations come to the same thing algebraically. But there is a fundamental difference between the analyses that underlie the two: "The central question in . . . velocity analysis was *how rapidly money is spent*. The central question in cash-balances analysis is why holders of cash *haven't spent it yet*."[18] D. H. Robertson expressed this distinction colorfully as money

[17] The $MV = PY$ formulation is also known as the Fisher equation, after Irving Fisher. More correctly, it is an adaptation of the original Fisher equation, $MV = PT$, in which T included all monetary transactions and not only final product, or Y. The $M = k(PY)$ formulation is known as the Marshallian equation, after Alfred Marshall. It is also popularly known as the Cambridge equation after the university at which Marshall, A. C. Pigou, D. H. Robertson, and Keynes (before the *General Theory*) developed this approach.

[18] A. G. Hart, P. B. Kenen, and A. D. Entine, *Money, Debt and Economic Activity*, 4th ed., Prentice-Hall, 1969, p. 216.

"on the wing" and money "sitting."[19] Money sitting leads one to ask why it is sitting and so, more specifically, why people demand money.

If $V = 4$, then $k = 1/4$. The former expression means that each dollar in the money supply is, on the average, used four times during the period for the purchase of final product. The latter means that people want to "keep by them" money equivalent in value to a certain stock of real goods and services, in this case an amount equal to one-fourth of the economy's output for the period.[20] The concept of the demand for money here appears explicitly. To illustrate, let us take an initial equilibrium situation in which people's actual money holdings are equal to desired money holdings, and then let us assume an increase in the money supply. As a result of the increase, the public is holding more money than it wants to hold at the existing price level, and it attempts to rid itself of this excess money by spending. On the assumption that output is at its full-employment level, the additional spending forces the price level up until the actual, larger money holdings become desired money holdings at the higher price level of output. To hold in money an amount sufficient to command one-fourth of the period's output requires more money at a higher price level for output; thus the rise in P with given Y equates the demand for M with the enlarged supply of M. A decrease in the money supply brings about a similar adjustment through a fall in the price level.

The basic assumption here is the same as in the velocity formulation of the simple quantity theory covered in the preceding chapter—people hold money for transactions purposes. Earlier in the chapter, the transactions demand for money in nominal terms was designated as M_t. Therefore, we had

$$M_t = k(PY)$$

and, assuming here that k is a constant, we may also write

$$M_t = P \cdot k(Y)$$

If nominal balances are demanded only for transactions purposes, i.e., $M_d = M_t$, then the total money supply, M_s, is absorbed in M_t and equilibrium between the supply of and demand for money is given by

$$M_s = P \cdot k(Y)$$

Simple Quantity Theory and Keynesian Theory

Like the quantity theorists, Keynes believed that the transactions demand for money was dependent on the level of income and not in any specific way on the interest rate, so that Keynes has $M_t = P \cdot k(Y)$. However, by adding the speculative demand for money, $M_{sp} = P \cdot h(r)$, Keynes denied that $M_d = M_t$. Therefore, he rejected the proposition that equilibrium between M_s and M_d can be stated as $M_s = P \cdot k(Y)$. As a result, Keynes's theory becomes far different from the quantity theory in the conclusions to which it leads.[21] Because equilibrium between the supply of and demand for money now occurs when

$$M_s = P \cdot k(Y) + P \cdot h(r)$$

a rise in M_s may in this case be absorbed in part by speculative demand. To show how an increase in M_s will be divided between the two demands requires the apparatus that will be developed in the following chapter. However, we can see here that a rise in M_s will no longer mean a rise in $P \cdot k(Y)$ equal to 4 times the rise in M_s, or with k being constant it will no longer mean a rise in PY equal to 4 times the rise in M_s,

[19] D. H. Robertson, *Money*, Univ. of Chicago Press, 1959, Ch. 2.

[20] See A. Marshall, *Money, Credit, and Commerce*, Macmillan, 1923, pp. 44–45.

[21] In the *General Theory*, Keynes implied that the demand for nominal transactions (and precautionary) balances varies with the price level but that the demand for nominal speculative balances does not. (See the *General Theory*, p. 199.) As noted earlier, post-Keynesian writers generally maintain that the speculative demand also varies with the price level and that is the treatment followed here.

assuming that the public wants to hold transaction balances equal to $(1/4)(PY)$. In Keynesian theory, an increase in M_s will affect PY only to the extent that it reduces r and to the extent that the reduction in r raises investment spending (and possibly consumption spending). Given a less than perfectly elastic speculative demand function, the increase in M_s will lead to some fall in r, and, given a less than perfectly inelastic MEC schedule, the fall in r will lead to some rise in investment spending and therefore some rise in PY. But the ratio of the rise in PY to the rise in M_s may be 0.4, 2.1, 3.8, or any other figure less than 4.[22] In short, changes in PY are no longer determinable from changes in M_s alone, as is true when transactions demand is the only demand for money.

The importance of this conclusion cannot be over-emphasized. It is one of the key differences between classical and Keynesian theory. Because the addition of the speculative demand for money to the quantity theory's demand for money equation rules out what otherwise is proportionality between changes in M_s and changes in PY, it means that one can no longer satisfactorily explain changes in aggregate demand, and therefore in PY, through changes in M_s. This calls for an alternative theory of aggregate demand and income determination—one that Keynes supplied. When we look at it in this way, we see the great importance of his addition of a speculative demand for money to the transactions demand that was the only demand recognized by the simple quantity theory.

[22]For a crude numerical illustration, assuming an initial equilibrium between the supply of and demand for money and that k is equal to 1/4, a rise in M_s of \$10 could then lead to a rise in PY of various possible amounts. If the speculative demand for money is very inelastic and the MEI curve is very elastic, a possible outcome could be $\Delta(PY)$ equal to \$38 with $k[\Delta(PY)]$ equal to \$9.50 and $h(\Delta r)$ equal to \$0.50. The ratio of the rise in PY to the rise in M is then 3.8. On the other hand, with the opposite combination of elasticities, a possible outcome could be $\Delta(PY)$ equal to \$4 with $k[\Delta(PY)]$ equal to \$1 and $h(\Delta r)$ equal to \$9. The ratio of the rise in PY to the rise in M in this case is 0.4.

Although Keynes treated the transactions (and precautionary) demand as interest inelastic, we saw that it actually becomes interest elastic at sufficiently high interest rates. Although Keynes made the speculative demand for money a function of the interest rate only, there are conditions under which it may also become a function of the level of income. If the transactions, precautionary, and speculative demands all depend to some extent on both the level of income and the interest rate, the separate demand functions for M_t and M_{sp} may be combined into one and written as

$$M_d = P \cdot f(Y,r)$$

This consolidated demand function avoids the artificiality of separating the demand for money into the three parts set forth by Keynes. People obviously do not divide their money holdings into three or more neat compartments to satisfy each of the three or more motives for holding cash. Statistically, for the economy as a whole, there is no way of separating the total money supply into active and idle balances without relying on very arbitrary assumptions. The consolidated equation, though general, has the advantage of showing that the demand for money for all purposes is a function of income level and interest rate.

Modern Quantity Theory

Expressed in its most basic form, the simple quantity theory of money makes the demand for nominal money balances depend only on the nominal income level or $M_d = P \cdot f(Y)$; the Keynesian theory adds the interest rate as a determinant to give us a different function: $M_d = P \cdot f(Y, r)$. During the post-Keynesian period, another theory was developed, which has come to be known as the new or modern quantity theory. It is held by its creator, Milton Friedman, to have its closest links with the old-fashioned quantity theory, although it clearly involves

some notable differences. Some other economists, however, find Professor Friedman's quantity theory in many ways closer to the Keynesian theory than to the old quantity theory that Keynes sought to displace.[23]

Although it is possible to distinguish clearly between the simple quantity theory and the Keynesian theory without great difficulty, the same is not true of the modern quantity theory vis-à-vis these other theories. The fact that Friedman sees the modern quantity theory as a restatement of the old one whereas others see it as an elaborate statement of Keynesian theory attests to the complexities present in it. We will not go into these complexities here but will note only some of the essentials needed to provide a perspective on how the modern quantity theory is related to the old-fashioned or simple quantity theory and to the Keynesian theory that we have already examined.[24]

As we saw above, one way of expressing the demand for money in the simple quantity theory is $M_d = k(PY)$, where k is a constant, and this equation may now be used as the point of departure for a sketch of some elements in Friedman's restatement of the quantity theory. This is not to imply that this equation correctly expresses Friedman's theory of the demand for money. He insists that the demand for money be treated as a demand for real balances just as a consumer's demand is a demand for real consumer goods; he makes use of permanent income rather than current income; he does not hold that k is a constant; and he introduces other differences. Several of these will be considered below, but the demand for money equation of the simple quantity theory is not so different from Friedman's to preclude its use for the limited purposes of this sketch.[25]

It has been noted that the simple quantity theory held k to be stable so that exact proportionality existed between M_s and PY. Furthermore, the simple quantity theory was a part of classical theory in which it was argued that full employment was the normal position of the economy so that in the short run Y was fixed at whatever output a fully employed economy was capable of producing. This meant that the rigid link between M_s and PY became a rigid link between M_s and P, or that the quantity theory was a theory of the price level. Like Keynesian theory, the new or modern quantity theory recognizes that output may be below its full-employment level other than temporarily so that Y is a variable in the short run and also that k in the simple quantity theory is a variable so that changes in M_s do not necessarily lead to proportional changes in PY. With k a variable, the value of PY will vary not only due to changes in M_s but also due to changes in k. However, while

[23]This is the convincing contention made by Don Patinkin in "The Chicago Tradition, the Quantity Theory, and Friedman," in *Journal of Money, Credit and Banking*, Feb. 1969, pp. 46–70. In Patinkin's words, "Milton Friedman provided us in 1956 with a most elegant and sophisticated statement of modern Keynesian theory—misleadingly entitled 'The Quantity Theory of Money—A Restatement.'" In a 1964 article, Friedman had practically said as much in the following words: "A more fundamental and more basic development in monetary theory has been the reformulation of the quantity theory of money in a way much influenced by the Keynesian liquidity preference analysis." "Postwar Trends in Monetary Theory and Policy," in *National Banking Review*, Sept. 1964, p. 4.

[24]Professor Friedman's own formulation is found in his short (but difficult) essay, "The Quantity Theory of Money—A Restatement," in M. Friedman, ed., *Studies in the Quantity Theory of Money*, Univ. of Chicago Press, 1956, pp. 3–21, reprinted in M. Friedman, *The Optimum Quantity of Money and Other Essays*, Aldine, 1969. For a longer and more readable presentation that also goes into the complexities, see S. Rousseas, *Monetary Theory*, Knopf, 1972, Chs. 9 and 10.

[25]This needs further qualification. We here refer to the equation that Friedman maintains gives a satisfactory explanation of the demand for money when applied to the aggregate time series data. This equation, used in his empirical work, is very different from the far more complex equation used to express his theoretical formulation of the demand for money. However, the latter equation involves variables like wealth and the ratio of nonhuman to human wealth on which there are either inadequate data or no data at all and is therefore not directly testable. It is approximated by one that resembles $M_d = k(PY)$ and can be tested.

the modern quantity theory recognizes that k is not stable in the sense of being some unvarying numerical value like 1/3 or 1/4, it argues that k is a stable *function* of a limited number of other variables that determine it. This means that although k will vary from time to time, its variations are not arbitrary but are explainable in terms of these other variables. In other words, it is the function determining k that is stable and not the value of k itself. Thus, unlike the old-fashioned quantity theory, the modern quantity theory does not hold that PY may be predicted from M_s alone but rather that PY may be predicted from M_s and from knowledge of the several variables that determine k.

However, in order that PY may be predicted in this way, it is necessary not only that k be a stable function of a limited number of variables but also that, under ordinary conditions, the elasticity of the demand for money with respect to these variables not be great. If it does exist, such a stable function in itself says only that k will not fluctuate unpredictably over time. Or, positively expressed, it says only that k will change by some specific amount in response to a change of some specific amount in any of the variables determining it.

In the simplified version of Friedman's theory, the two major variables determining k are the interest rate and the rate of change in the price level. In the simple quantity theory, M_d/P is equal to M_t/P because that theory holds that the only demand for money is the transactions demand. The equation in that case is simply $M_d/P = k(Y)$ in which k is a constant determined by certain institutional and structural characteristics of the economy that ordinarily do not change significantly in the short run. The demand for money equation in the version of Friedman's theory being presented here is

$$M_d/P = k(r, \dot{P}) \cdot Y$$

in which $\dot{P}$ is the rate of change of the price level. That is to say, the quantity of real balances the public wants to hold is no longer a fixed amount at any level of real income but an amount that varies at that level of income with the interest rate and the rate of change of the price level.

It is the function, $k = k(r, \dot{P})$, in this equation which Friedman says is stable. Accepting that this function is indeed stable, it still remains that the closeness of the tie between the demand for real balances and real income will be less, the greater the elasticity of demand for real balances with respect to r and $\dot{P}$. If the demand for money is highly elastic with respect to r, then relatively small changes in r will produce relatively larger changes in M_d/P. Given the fact that the interest rate does change frequently and sometimes appreciably and that M_d/P may therefore change frequently and sometimes appreciably, the result may be that *there is no longer a stable relationship between M_d/P and Y in the equation, $M_d/P = k(r, \dot{P}) \cdot Y$, although k itself may be a perfectly stable function of r and $\dot{P}$ which determine it.* Thus, in order to satisfy the modern quantity theory contention that real balances are a stable function of real income, not only must the k function be stable but the relevant elasticities must also be sufficiently small. There are conditions under which the modern quantity theory predictions will hold even with high elasticities, but it does not appear that these conditions generally prevail.

However, as Friedman reads the empirical record, the problem here noted does not arise in practice because the relevant elasticities are found to be so small as to rule out any significant influence on M_d/P. In the case of the elasticity with respect to the rate of interest (or, in Friedman's complete formulation, several rates of return—those on money, bonds, and equities), Friedman grants that economic theory would lead one to expect that interest rates are one of the variables influencing the public's desired money balances and thus affecting k. In this regard his theoretical formulation of the quantity theory closely resembles Keynesian theory with its inclusion of the interest rate in the demand-for-money equation. But what one would expect from theory is not, according to

Friedman, what one actually finds in the empirical record. As he interprets that record, the influence of the interest rate on the demand for money, though present, is quite minor.[26] In other words, he holds that the interest rate need not be included in a demand-for-money equation in order for that equation to provide a satisfactory explanation of the demand for money. In contrast, all other empirical investigations find that the interest rate has more than the minor influence on the demand for money that Friedman has found.[27] These studies uniformly find interest elasticities significantly different from zero and of a magnitude capable of producing instability in the size of k. To the degree that the results of these studies are the more correct—and the degree of uniformity in their results suggests that they are—they represent a damaging blow to Friedman's quantity theory.

The second major variable affecting k is the percentage change in the price level. This variable and the interest rate together determine the cost of holding money, the interest rate reflecting the rate of return one forgoes by holding money and the rate of change in the price level reflecting the rate of change in the real value of the money one holds. The rate of change in the price level must be carefully distinguished from alternative absolute price levels. As the demand for money is expressed here as a demand for real balances, the influence of the absolute price level is already reflected in M_d/P. Other things being equal, a higher P will mean a correspondingly higher M_d or no change in the demand for real balances. The absolute level of P does not affect k. However, as prices are in the process of rising, the rate at which prices are rising, i.e., the size of $\dot{P}$, does alter the amount of real balances demanded. The greater is $\dot{P}$, the greater is the rate at which the real value of each dollar held shrinks or the more rapid is the rise in the cost of holding money. Accordingly, the greater is $\dot{P}$, the greater is the decline in k.

In his empirical work, Friedman was unable to establish that the percentage rate of change of prices had any influence at all on k and through k on the demand for money, although a historical analysis persuades him that such an effect is present. When he combines this empirical finding with his empirical finding that the interest rate does not appreciably affect the demand for money, his conclusion is that changes in the cost of holding money do not significantly affect the amount of money held. Because the two variables that measure the cost of holding money are the principal variables that determine the size of k, Friedman's contention that there is no significant elasticity in the demand for money with respect to these variables brings him to the modern quantity theory conclusion that k, although not stable in itself, varies sufficiently little to make for a stable relationship between the demand for real money balances and the level of real income.

As was noted above, the empirical record as read by other investigators reveals a sufficiently high interest elasticity of demand to make Friedman's position unacceptable. As they see it, the variations in k caused by an interest-elastic demand for money are too large and uncertain to yield the conclusions the modern quantity theory claims. For this reason, they are led to a demand-for-money equation of the Keynesian type like $M_d/P = k(Y) + h(r)$ instead of a demand-for-money equation like $M_d/P = k(y)$ which is essentially what emerges from the

[26] See his "The Demand for Money: Some Theoretical and Empirical Results," in *Journal of Political Economy*, Aug. 1959, pp. 327–51, reprinted in R. A. Gordon and L. R. Klein, eds., *Readings in Business Cycles*, Irwin, 1965, pp. 427–55.

[27] For further discussion of this question and a review of the empirical evidence, see J. T. Boorman, "The Evidence on the Demand for Money: Theoretical Formulations and Empirical Results," in J. T. Boorman and T. M. Havrilesky, *Money Supply, Money Demand, and Macroeconomic Models*, Allyn and Bacon, 1972, pp. 248–86. As shown by the table on p. 266, single equation studies employing a long-term rate of interest find elasticities in the range from −0.4 to −0.9 and those employing a short-term rate of interest find elasticities in the range from −0.1 to −0.5. These studies are unanimous in concluding that the rate of interest is an important factor in explaining variations in the demand for money.

modern quantity theory equation, $M_d/P = k(r, P) \cdot Y$, after denying any significant role to r and P as influences on M_d/P.

The particular aspect of the modern quantity theory that we have reviewed in the preceding paragraphs is the one that is most widely known and probably the one of greatest practical importance. It is this aspect that is most closely related to monetarism and that doctrine's contention that the supply of money is overwhelmingly important in determining the level of income. Or, in nominal terms, $PY = (1/k) \cdot M_s = V \cdot M_s$. It is also this aspect that leads to the rejection of the Keynesian emphasis on government budgets or fiscal policy as a force affecting the level of income. A more complete treatment of the modern quantity theory would take us more deeply into this aspect as well as into a number of other aspects not mentioned here. We will complete the present sketch of the modern quantity theory with a brief look at just one of its other important aspects.

The theoretical formulation of the modern quantity theory contains the economy's wealth (broadly conceived to include human as well as nonhuman capital) rather than the level of income as the "scale" variable in the demand for money. This reflects a difference in view as to the role of money in the economy. If that role is viewed quite narrowly—that money is held merely as a medium of exchange, as something with which to carry out transactions—then the income variable would appear to be the more appropriate scale variable. This, of course, follows from what is probably a close correspondence between the volume of transactions and the level of income. However, if the role of money is seen as something larger, wealth turns out to be the more appropriate scale variable. In this broader view taken by Friedman, money becomes one among the various productive assets held by firms and individuals. To the firm, money is a capital good much as the tools and machines used in the productive process are capital goods. The demand for money is comparable to the demand for capital—each is an input and each is held for the services that it yields. To the individual, money is an asset that provides utility like other assets. He compares the utility derived from the asset money with that from durable goods or with that from the return on bonds and other forms of wealth. All of this does not deny the role of money as a medium of exchange; it simply takes that for granted and sees money as an asset that yields services beyond the mere mediation of transactions. With money viewed as an asset in this way, the amount that the public wants to hold may be expected to increase as the public's total assets or wealth grows larger over time. The demand for money thus becomes a part of the larger problem of the demand for all kinds of wealth, financial and nonfinancial, human and nonhuman.

When economists seek to investigate empirically this concept of the demand for money, they face the problem of inadequate data on wealth. Human capital, in particular, is most difficult to quantify. In his own empirical work, Friedman employs the concept of permanent income as a proxy for the stock of wealth. His definitions are such that permanent income is the appropriately discounted sum of the expected income flows generated by the various assets that make up total wealth, human and nonhuman. As we saw in Chapter 8 in examining the concept of permanent income, it too is not directly measurable but may be approximated in the way described there. The upshot of this is that in the modern quantity-theory equation real income or Y is understood to be permanent income and not current income as in the simple quantity theory and in the Keynesian-theory equations. Furthermore, Y is there only as a proxy for wealth, which is not readily measurable. This, then, is another of the differences between the modern quantity theory and earlier theory: its broader concept of the role of money in the system and, consistent with this, its use of a comprehensive concept of wealth as the appropriate scale variable in the demand-for-money equation.

Differences between the modern quantity theory and earlier theory in such matters as the appropriate scale variable in the demand-for-money equation and the elasticity of the demand for money with respect to the cost of holding money may be expected to diminish as empirical work in the area proceeds. Friedman has held that the basic differences among economists are empirical and not theoretical. His reading of the historical record has brought him to a demand-for-money equation that, much like the old-fashioned quantity-theory equation, makes changes in the money supply the element of overwhelming importance in explaining changes in the level of money income. Other economists read the empirical evidence differently and in so doing find no such dominant role for money in the explanation of changes in the level of income. To repeat an elementary but basic point, most economists who have done empirical work in this area have found that the demand for money is significantly influenced by the interest rate, and this finding leads one right back to the Keynesian theory of the demand for money and thereby away from the very close link between changes in money and changes in the income level postulated by the quantity theory.

Over the years ahead, more detailed and more subtle examination of the statistical record can be expected to produce evidence supporting one side or the other in this major debate between the quantity and Keynesian theories of the demand for money. We will not pursue that debate further. However, for purposes of our work in the chapters ahead, we must have a theory of the demand for money, and we accept the Keynesian theory that was developed in the first part of this chapter. That theory is described by the general equation $m_d = M_d/P = f(Y, r)$, but we will find it convenient in what follows to identify a transactions demand that depends only on Y and a speculative demand that depends only on r. This returns us to the equation in which we dichotomize the total demand:

$$m_d = \frac{M_d}{P} = k(Y) + h(r)$$

With this as our theory of the demand for money and with the liquidity preference theory of the interest rate that is derived from it, we can now proceed to combine the theory of money and interest developed earlier in this chapter with the theory of income determination developed in earlier chapters. This will provide us with a model which simultaneously determines the equilibrium level of income and the equilibrium interest rate.

chapter 16

The Extended Model: Fixed Price Level

In previous chapters we developed separately the theory of income determination and the theory of money and interest. Although this procedure provided an orderly introduction to the relevant theory, it must now be recognized for the simplification that it is—the two parts are actually so related that what happens in one depends on what happens in the other. In developing the Keynesian theory of income determination, we found that a rise in investment spending would raise the equilibrium level of income by an amount equal to the multiplier times the rise in investment spending. However, we implicitly assumed that the interest rate was given. If we now admit the interest rate as a variable in the system, the rise in investment spending will, by raising the level of income, also force up the interest rate. This in turn will discourage investment, and the actual rise in the equilibrium level of income will be less than it would otherwise be. Similarly, in developing the theory of money and interest in Chapter 15, we saw that an increase in the money supply would reduce the interest rate, as shown by the movement down the given demand curve for money. However, this curve assumed a given level of income. If we now admit the income level as a variable in the system, the increase in the money supply will, by lowering the interest rate, stimulate investment spending and raise the level of income. This will increase the transactions demand for money, and the actual fall in the interest rate will be less than it would otherwise be. Thus, it would seem that the interest rate and the level of income are linked in a complicated manner. In this and the following two chapters, we will construct and employ an extended model that can accommodate this and other complications.[1]

Our procedure is as follows. This chapter constructs a model which, like those in Part Two, is based on the assumptions that the ag-

[1]The construction here will be almost entirely graphic. For an algebraic formulation of the same elementary model covered in this chapter, see the Appendix to Chapter 16 of E. Shapiro, *Macroeconomic Analysis—A Student Workbook*, 4th ed., Harcourt Brace Jovanovich, 1978. A concise algebraic treatment of a less elementary *IS–LM* model is provided in W. L. Smith and R. L. Teigen, *Readings in Money, National Income, and Stabilization Policy*, 3rd ed., Irwin, 1974, pp. 1–22. The model, including a variable price level, is developed in R. S. Holbrook, "The Interest Rate, Price Level, and Aggregate Output," in the same volume, pp. 38–60.

gregate supply curve is perfectly elastic up to the full employment level of output, that the function does not shift, and that the level of output at which the economy operates varies along the range below the full employment level. Therefore, so far as the price level is concerned, changes in aggregate demand can affect only the output level and not the price level. There is a fixed price level which is the basis for the title of the chapter: The Extended Model: Fixed Price Level. Among other things, the following chapter replaces the perfectly elastic aggregate supply curve with an upward-sloping supply curve. As a result, shifts in the aggregate demand curve may affect the price level as well as the output level. The price level is variable and the chapter is entitled The Extended Model: Variable Price Level. Both of these chapters, Chapters 16 and 17, deal with a closed economy. Chapter 18, the third of this set of chapters, extends the analysis to the open economy and introduces the balance of payments into the analysis. We there have the title: The Extended Model: Foreign Sector Included.

THE GOODS MARKET AND THE MONEY MARKET

Our model consists of two parts: The first draws together the determinants of equilibrium in the market for goods, and the second draws together the determinants of equilibrium in the market for money. For a two-sector economy, we found in Chapter 4 that goods market equilibrium is found at that level of Y at which the sum of $C + I$ generated by that level of Y is just equal to that level of Y. Goods market equilibrium is also defined by an equality between saving and investment. At the level of Y at which $S = I$, the leakage from the income stream into S is exactly matched by offsetting I. In both approaches we have the result that the equilibrium level of income or output is that at which aggregate supply of goods and aggregate demand for goods are equal or $AS = AD$. Money market equilibrium is defined by an equality between the supply of and the demand for money—$m_s = m_d$—the condition that gave us the equilibrium interest rate. Or at the interest rate at which $m_s = m_d$, there is money market equilibrium.[2]

The particular level of income at which there is goods market equilibrium depends in part on conditions in the money market, and the particular interest rate at which there is money market equilibrium depends in part on conditions in the goods market. For a preliminary look at what is involved, let us briefly review the simplest possible Keynesian model as shown in Figure 16-1. Given the $C + I_1$ curve in Part A and the S and I_1 curves in Part B, the equilibrium level of Y is Y_1, at which level $AS = AD$, in Part C. If investment depends at all on the interest rate, the $C + I_1$ curve in Part A and the I_1 curve in Part B must have been drawn on the assumption of some particular interest rate. A lower interest rate, other things being equal, would indicate a different position for the $C + I$ curve, say $C + I_2$ instead of $C + I_1$, and a different position for the I curve, say I_2 instead of I_1. This,

[2]A more complete general equilibrium model will also include the market for factors of production, which, because of the short-run assumption of a fixed capital stock, becomes the market for labor. Equilibrium in this market requires equality between the supply of and the demand for labor. From a Keynesian viewpoint, disequilibrium in this market in the form of an excess supply of labor—that is, unemployment—can be corrected by policies designed to raise the level of output—that is, to shift the equilibrium in the goods market to a higher level of goods output whose production in turn calls for employment of more labor. From a classical viewpoint, the same disequilibrium would be removed automatically by falling wages and prices in a system characterized by such flexibility. Following the development of the basic model, which is limited to the goods and money markets, attention will be given to these other questions in the following chapter.

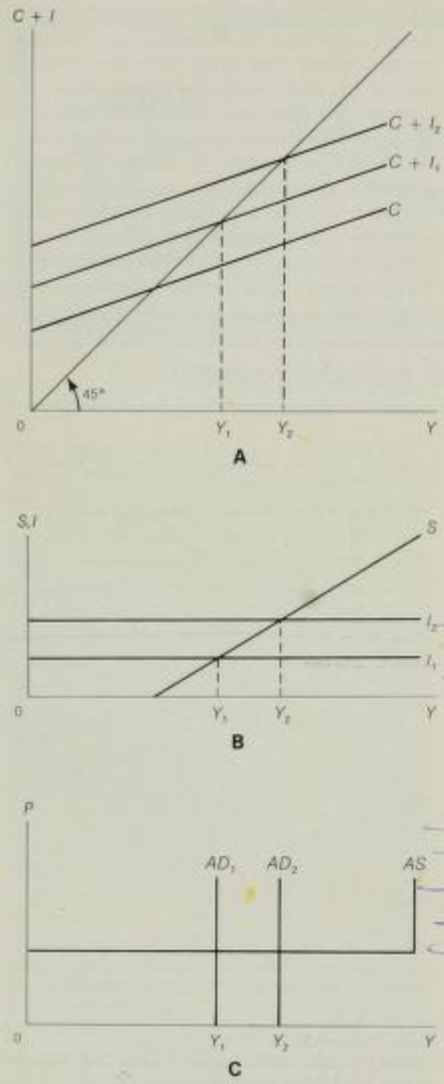


FIGURE 16-1
Equilibrium levels of income

in turn, would indicate a different equilibrium income level, Y_2 instead of Y_1. In Part C, *AD* would shift from AD_1 to AD_2 and intersect *AS* at Y_2. Figure 16-1, however, does not tell us what the interest rate may be—it assumes some rate and proceeds from there.

Figure 16-2 shows the determination of the equilibrium interest rate. Given the m_s and m_{d_1} curves, the equilibrium rate is r_1, at which the demand for and the supply of money are equal, or $m_{t_1} + m_{sp_1} = m_{d_1} = m_s$. However, because the demand for money is composed in part of the transactions demand which depends on the level of income, the m_{d_1} curve must have been drawn on the basis of some assumed income level that defined m_{t_1}. A higher income level, other things being equal, would indicate a different position for the curve, say m_{d_2} instead of m_{d_1}. This would indicate a different equilibrium rate of interest, r_2, at which $m_{t_2} + m_{sp_2} = m_{d_2} = m_s$. Figure 16-2, however, does not tell us what the level of income may be—it assumes some income level and proceeds from there.

It appears that we cannot determine the equilibrium income level without first knowing the interest rate and that we cannot determine the equilibrium interest rate without first knowing the income level. Somehow *Y* and *r* must be determined simultaneously. Although this cannot be done through Figures 16-1 and 16-2, there are nonetheless a particular income level and interest rate that simultaneously provide equilibrium in the goods market behind Figure 16-1 and in the money market behind Figure 16-2. The model to be developed in this chapter provides this simultaneous solution of the two equilibrium values and sheds light on some other important problems and policy questions.[3]

[3]This model was originally developed by J. R. Hicks in his article "Mr. Keynes and the 'Classics': A Suggested Interpretation," in *Econometrica*, April 1937, pp. 147–59, reprinted in W. Fellner and B. F. Haley, eds., *Readings in the Theory of Income Distribution*, Irwin, 1946, pp. 461–76. See also F. Modigliani, "Liquidity Preference and the Theory of Interest and Money," in F. A. Lutz and L. W. Mints, eds., *Readings in Monetary Theory*, Irwin, 1951, particularly pp. 190–206.

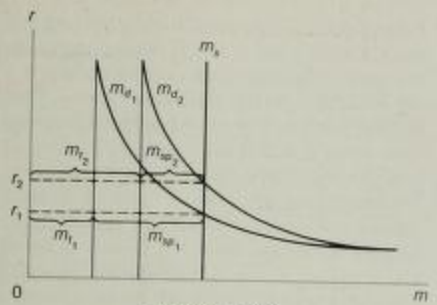


FIGURE 16-2
Equilibrium levels of the interest rate

Equilibrium in the Goods Market

Since equilibrium in the goods market requires that $Y = C + I$ and $S = I$, all the factors that cause the consumption function and therefore the saving function to shift and all the factors that cause the investment function to shift influence the determination of this equilibrium. Although other factors may be introduced once the basic model is developed, we assume here that investment is a function of the interest rate alone and that consumption and therefore saving is a function of income alone. From the $C + I$ approach, we then have, in general terms, the following three equations to cover the goods market.

Consumption function: $C = C(Y)$
Investment function: $I = I(r)$
Equilibrium condition: $Y = C(Y) + I(r)$

From the S, I approach, we have, in general terms, the following three equations to cover the goods market.

Saving function: $S = S(Y)$
Investment function: $I = I(r)$
Equilibrium condition: $S(Y) = I(r)$

One may develop the diagrammatic analysis that follows on the basis of either or both of the approaches, but attention here will be limited to that based on the S, I approach.

The set of equations for that approach may be shown graphically as in Figure 16-3. Part A gives the MEI (investment spending) schedule, showing that investment spending varies inversely with the interest rate. The straight line in Part B is drawn at a 45° angle from the origin. Whatever the amount of planned investment measured along the horizontal axis of Part B, equilibrium requires that planned saving measured along the vertical axis of Part B be the same. Thus, all points along the 45° line in Part B indicate equality of saving and investment. Part C brings in the saving function, showing that saving varies directly with income.

The IS curve in Part D is derived from the other parts of the figure. To illustrate, let us assume an interest rate of 3 percent in Part A, indicating that investment is $20 per time period.[4] In Part B, to satisfy the equality between S and I, saving must also be $20, as shown on the vertical axis. In Part C, we find that saving will be $20 only at an income level of $120.[5] Finally, bringing together Y of $120 from Part C and r of 3 percent from Part A, we have one combination of Y and r at which $S = I$ (and $Y = C + I$). If we assume the lower interest rate of 2.5 percent, Part A indicates that investment will be $30, which gives us an income level of $140 in Part C. Therefore, Y of $140 and r of 2.5 percent is another combination of Y and r at which $S = I$. Other combinations could be found in the same way by starting with other assumed interest rates and finding the income level at which saving is equal to the amount of investment indicated by that interest rate. Connecting these combinations gives us the IS curve in Part D.[6]

[4] All dollar amounts are in billions.

[5] The saving function $S = S_a + sY$ is here $S = -40 + \frac{1}{2}Y$.

[6] Alternatively, the combinations of Y and r at which $S = I$ could just as well be determined graphically by starting with assumed levels of Y. Thus, assuming Y of $120, Part C shows that S will be $20. Moving from Part C through Part B to Part A, one finds that I of $20 is consistent with r of 3 percent. Therefore, in Part D, Y of $120 and r of 3 percent identifies one combination at which $S = I$.

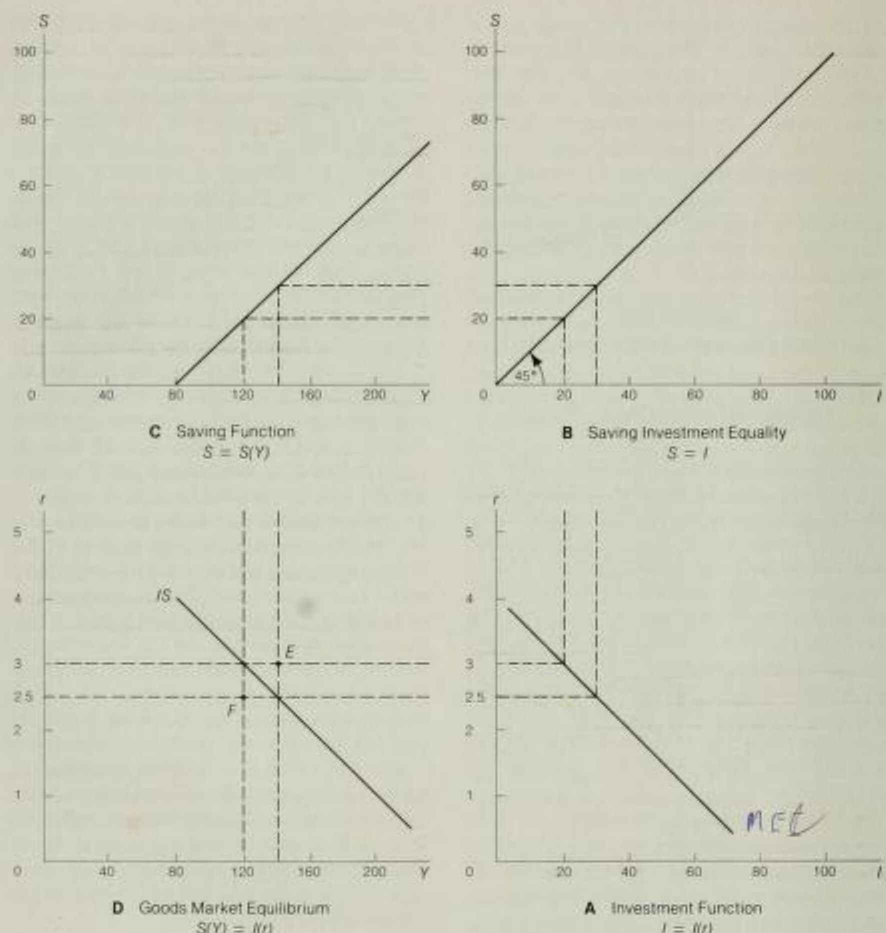


FIGURE 16-3
Goods market equilibrium

We find that there is no longer a single level of income at which $S = I$ but a different level for each different interest rate. The higher the interest rate, the lower the level of income at which $S = I$. Viewed in one way, this follows from the fact that a high r means a low I, and a low I, through the multiplier, means a low Y. Viewed in another way, it follows from the

fact that a low Y means a low S. Since equilibrium requires that $S = I$, a low S means a low I, and a low I is the result of a high r. Although the *IS* function indicates that equilibrium in the goods market will be found at a lower level of income for a higher interest rate, it alone does not tell us what particular combination of Y and r will be found in any specific time period. All combinations on the *IS* function are equally possible equilibrium combinations of Y and r in the goods market.

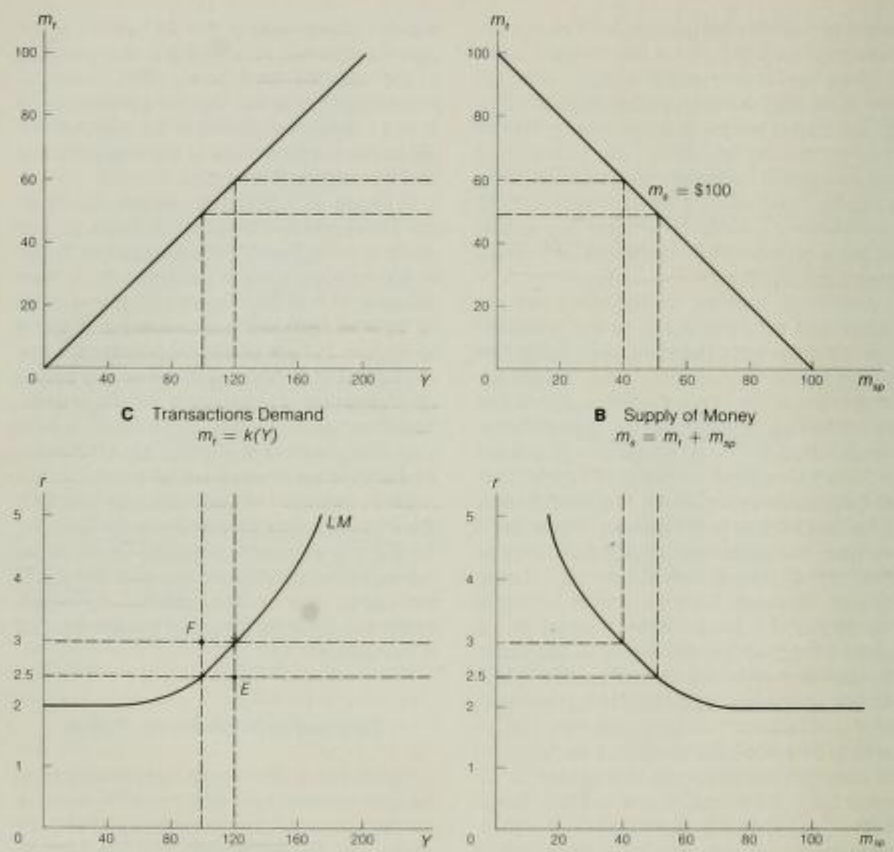
Identifying all those combinations that are equilibrium combinations does not, however, mean that the actual combination in each time period will be one of them. There may be disequilibrium in the goods market. Suppose that the actual combination is the disequilibrium combination of Y = \$140 and r = 3 percent indicated as point E in Part D of Figure 16-3. At the income level of \$140, S will equal I only if the interest rate is 2.5 percent. Therefore, if we have the combination of this \$140 income level with an interest rate of 3 percent, S must exceed I because I will be smaller at a rate higher than 2.5 percent than it will be at 2.5 percent but S will be unchanged. In the model, S depends only on the level of income, which is here unchanged at \$140. The combination of Y = \$140 and r = 3 percent may also be seen to be a disequilibrium from a second point of view. At the interest rate of 3 percent, S will equal I only if the income level is \$120. Therefore, if we have the combination of this 3-percent interest rate with an income level of \$140 as at point E, S must exceed I because S will be larger at an income level above \$120 than it will be at the level of \$120 but I will be unchanged. In the model, I depends only on the interest rate, which is here unchanged at 3 percent. It follows that for any combination of Y and r located anywhere in the space to the *right* of the *IS* curve, the same conclusion may be drawn that was drawn for point E: there is a disequilibrium in which S exceeds I (and $Y > (C + I)$).

By the same line of reasoning, it may be seen that the combination of Y = \$120 and r = 2.5 percent indicated as point F is a disequilibrium of the opposite kind: here I must exceed S. Generalizing as above, for any combination of Y and r located anywhere in the space to the *left* of the *IS* curve, there is a disequilibrium in which I exceeds S (and $(C + I > Y)$).

In Figure 16-1, perfectly inelastic *AD* curves are positioned in Part C at the levels of Y at which $S = I$ in Part B. With two I curves in Part B, there are two levels of Y at which $S = I$. Now, however, in Part D of Figure 16-3 all the levels of Y shown below the *IS* curve are levels at which $S = I$. Again, given the assumption that the *AD* curve is perfectly inelastic and the *AS* curve is perfectly elastic over the range within which the equilibrium level of Y is found, is there a particular level of Y at which the *AD* curve is located? As the *IS* curve shows that $S = I$ at a different level of Y for each different level of r, the answer requires that we know the level of r. To find this we must turn to an analysis of the money market and beyond this to an analysis of the goods and money markets combined, which will reveal the particular interest rate that is being sought.

Equilibrium in the Money Market

Equilibrium in the money market requires an equality between the supply of and the demand for money. The Keynesian theory of the demand for money makes the transactions demand (which is here combined with the precautionary demand) a direct function of the income level alone, or $m_t = k(Y)$, and the speculative demand an inverse function of the interest rate alone, or $m_{sp} = h(r)$. Total demand for money is $m_d = m_t + m_{sp} = k(Y) + h(r)$. The supply of money m_s is determined outside the model—it is exogenous. This may be written $m_s = m_a$, in which m_a is simply the amount of money that exists, an amount determined by the monetary authorities. Of course, the monetary authorities determine only the nominal



C Transactions Demand $m_t = k(Y)$

B Supply of Money $m_s = m_t + m_{sp}$

D Money Market Equilibrium $m_s = k(Y) + h(r)$

A Speculative Demand $m_{sp} = h(r)$

FIGURE 16-4
Money market equilibrium

stock of money, M_s, but with P assumed to be stable, determination of M_s also determines m_s. This gives us three equations to cover the money market.

Demand for money: $m_d = k(Y) + h(r)$
Supply of money: $m_s = m_s$
Equilibrium condition: $m_d = m_s$

This set of equations is shown graphically in Figure 16-4. Part A shows the speculative demand for money as a function of r. Part B is drawn to show a total money supply of \$100, all of which must be held in either transactions balances or speculative balances. The points along the line indicate all the possible ways in

which the given money supply may be divided between m_s and m_{sp}. Part C shows the amount of money required for transactions purposes at each level of income on the assumption that $k = 1/2$. The *LM* curve of Part D is derived from the other parts as follows.

Assume in Part A an interest rate of 3 percent; at this interest rate the public will want to hold \$40 in speculative balances. In Part B, subtracting the \$40 of speculative balances from a total money supply of \$100 leaves \$60 of transactions balances, an amount consistent with an income level of \$120 as shown in Part C. Finally, in Part D, bringing together Y of \$120 from Part C and r of 3 percent from Part A, we have one combination of Y and r at which $m_d = m_s$ or at which there is equilibrium in the market for money. If we assume the lower interest rate of 2.5 percent, Part A indicates that speculative balances will be \$50, Part B indicates that transactions balances will be \$50, and Part C indicates the income level of \$100 as that consistent with transactions balances of \$50. This gives us another combination of Y and r—\$100 and 2.5 percent—at which $m_d = m_s$. Other such combinations can be determined in the same way. The function in Part D labeled *LM* results when these combinations are connected.[7]

Although particular characteristics of the *LM* function will call for attention later, in general it is seen that the function slopes upward to the right. With a given stock of money, money market equilibrium is found at combinations of high interest rates and high income levels or low interest rates and low income levels. Viewed in one way, this can be seen to follow from the fact that a high level of income calls for relatively large transactions balances, which, with a given money supply, can be drawn out of speculative balances only by pushing up the interest rate. Viewed in another way, it can be seen to follow from the fact that a high interest rate is one at which speculative balances will be low; this releases more of the money supply for transactions balances, which will be held in such balances only at a correspondingly high level of income. Although the *LM* function indicates in this fashion why equilibrium in the money market will occur at a higher interest rate for a higher level of income, it alone cannot tell us what particular combination of Y and r will be found in any given time period. All combinations on the *LM* function are equally possible equilibrium combinations in the money market.

As with the *IS* curve, identifying all combinations at which $m_s = m_d$ does not mean that the actual combination in each time period will be one of them. It may be one that involves a disequilibrium in the money market. To illustrate, consider the disequilibrium combination indicated by point E in Part D of Figure 16-4. At the income level of \$120, m_d will equal m_s only if r is 3 percent. Therefore, if we have the combination of this \$120 income level with an interest rate of 2.5 percent, m_d must exceed m_s because m_d will be larger at $r = 2.5$ percent than it will be at $r = 3$ percent. The quantity of money demanded for speculative purposes rises with a lower interest rate, while the total supply of money is fixed. Looking at this in an alternative way, starting with the interest rate of 2.5 percent, m_d will equal m_s only if the income level is \$100. Therefore, if we have the combination of this 2.5-percent interest rate with an income level of \$120 shown as point E, m_d must exceed m_s because m_d will be larger at an income level above \$100 than it will be at the income level of \$100. The quantity of money demanded for transactions rises with a higher income, while the total money supply is, as before, fixed. What has been concluded for the combination indicated by point E may be concluded for any combination located anywhere in the space to the right of the *LM* curve; any

[7] As with the *IS* curve, the combinations of Y and r at which $m_d = m_s$ could just as well be determined graphically by starting with assumed levels of Y. Thus, assuming Y of \$120, Part C shows that m_t will be \$60. Subtracting \$60 from the total money supply of \$100 leaves \$40. As Part A shows, this is an amount the public will be willing to hold in speculative balances, m_{sp}, at r of 3 percent. In Part D, Y of \$120 and r of 3 percent therefore identifies one combination of Y and r at which $m_d = m_s$.

combination of Y and r in this space is necessarily a disequilibrium combination in which m_d exceeds m_s.

Following the same kind of reasoning, one may see that the combination of Y = \$100 and r = 3 percent indicated as point F is a disequilibrium of the opposite kind: m_s must here exceed m_d. Generalizing as before, one may say that there is a disequilibrium in which m_s exceeds m_d for any combination of Y and r located anywhere in the space to the left of the LM curve.

TWO-MARKET EQUILIBRIUM—THE GOODS AND MONEY MARKETS

Equilibrium between the supply of and demand for goods is possible at all combinations of Y and r indicated by the IS curve; similarly, equilibrium between the supply of and demand for money is possible at all combinations of Y and r indicated by the LM curve. However, there is only one combination of Y and r at which both the supply of goods equals the demand for goods and the supply of money equals the demand for money and it is defined by the intersection of the IS and LM curves derived in Figures 16-3 and 16-4 and brought together in Figure 16-5. In this illustration, equilibrium in both markets occurs with Y = \$120 and r = 3 percent.

With the equilibrium rate of interest established at 3 percent, of all the levels of Y at which $S = I$, the one which determines the location of the AD curve in Part B of Figure 16-5 is the one consistent with the equilibrium interest rate of 3 percent. Hence, Part B shows the AD curve at Y = \$120 because only at this level of Y will $S = I$ with the interest rate given at 3 percent.

In Part A, every possible combination of Y and r other than that given by the intersection of the IS and LM curves is one at which there is disequilibrium in either the goods market or the money market or in both. All those combinations that do not lie either on the IS or on the LM curve fall into this last category. Since all such combinations do not lie on a line, they necessarily lie in one of the four areas identified by the Roman numerals I through IV. As we saw earlier, any combination of Y and r that

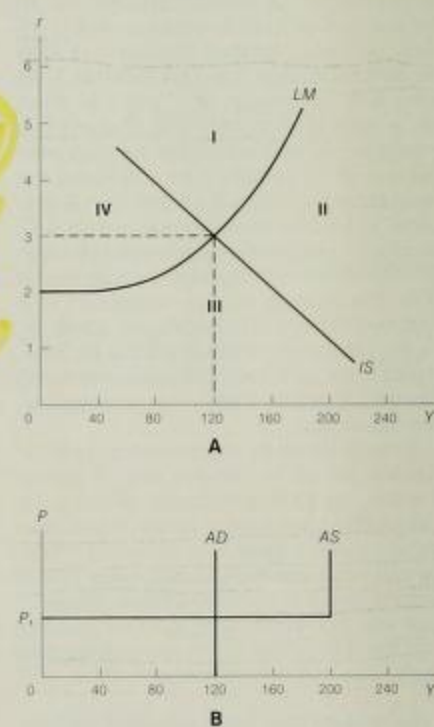


FIGURE 16-5
Equilibrium in the goods and money markets and the aggregate demand curve

lies anywhere to the right of the *IS* curve is a combination at which $S > I$ and $Y > (C + I)$, and the opposite is true for any combination of Y and r anywhere to the left of the *IS* curve. Similarly, any combination of Y and r anywhere to the right of the *LM* curve is a combination at which $m_d > m_s$ and the opposite is true for any combination to the left of the *LM* curve. Accordingly, each of the four spaces may be distinguished from the other three in terms of the relationships between the supply of and demand for goods and between the supply of and demand for money for any combination of Y and r that falls within that space. We find the following.

	GOODS MARKET	MONEY MARKET
In Space I:	$I < S, (C + I) < Y$	$m_d < m_s$
In Space II:	$I < S, (C + I) < Y$	$m_d > m_s$
In Space III:	$I > S, (C + I) > Y$	$m_d > m_s$
In Space IV:	$I > S, (C + I) > Y$	$m_d < m_s$

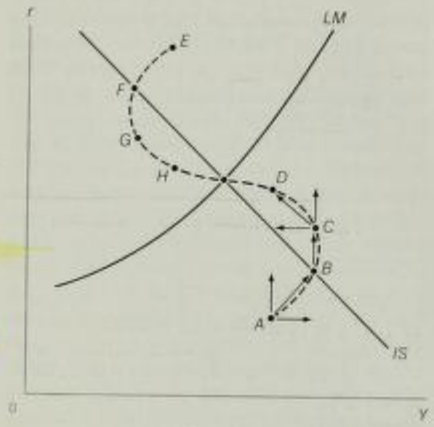


FIGURE 16-6
Possible paths of movement to equilibrium in the goods and money markets

From the analysis of the goods market considered in isolation, we know that a situation in which $I > S$ or $(C + I) > Y$ will lead to a rise in income and vice versa. From the analysis of the money market considered in isolation, we know that a situation in which $m_s > m_d$ will lead to a fall in the interest rate and vice versa. What we now have in the four spaces laid out in Figure 16-5 are various combinations of *IS* and *LM* disequilibrium situations. Simply knowing the direction in which the income level tends to move in response to an excess supply or excess demand for goods and the direction in which the interest rate tends to move in response to an excess supply or excess demand for money, it is possible to trace out in a nonrigorous fashion a possible path that the income level and the interest rate may follow in response to any given disequilibrium situation.

In Figure 16-6 we assume the economy to be located at the disequilibrium combination of Y and r indicated by A, which is in Space III. Here there is an excess demand for goods and an excess demand for money. The excess demand for goods tends to raise the income level as indicated by the horizontal arrow originating at A; the excess demand for money tends to push up the interest rate as indicated by the vertical arrow originating at A. With these forces at work, it is not unreasonable to expect the economy to move along the path designated by the arrow from A to B. Next with the economy at B, the supply of and demand for goods are equal, because we are on the *IS* curve. But we are still at a point to the right of the *LM* curve, so the demand for money exceeds the supply of money. Therefore, a force is at work to push up the interest rate and the next movement may be along the arrow from B to C. At C we still find an excess demand for money, which again tends to push up the interest rate as indicated by the vertical arrow originating at C. However, at C we find an excess supply of goods, which tends to reduce the income level as shown by the horizontal arrow originating at C. These forces may on balance cause the economy to move

along the path described by the arrow running from C to D. At D the forces are the same as at C, and the result again is a movement of the same kind. The combination of income level and interest rate may change in this way over time until finally the system reaches that one combination of Y and r at which both markets clear. Although the several discrete steps here traced out help one to see the underlying process at work, the actual process would be a continuous one in which Y and r might move along a path like that indicated by the dashed line running from A to D and then to the intersection of the two curves.

Instead of starting at A, we could start at any other disequilibrium point in Figure 16-6 and trace the movement of Y and r toward the single pair of equilibrium values in the same way. No matter what disequilibrium point one starts with, all one need do is (1) identify whether $I > S$, $I < S$, or $I = S$, or in terms of aggregate spending whether $(C + I) > Y$, $(C + I) < Y$, or $(C + I) = Y$, which tells whether Y will tend to rise, fall, or remain unchanged; (2) identify whether $m_d > m_s$, $m_d < m_s$, or $m_d = m_s$, which tells whether r will tend to rise, fall, or remain unchanged; and (3) establish the direction of movement of the Y, r combination that is indicated by the forces found to be at work in (1) and (2). For example, starting with any point in Space I such as E, we have a force tending to reduce the income level and a force tending to lower the interest rate. The reader may trace the discrete steps from E through H, which, it will be seen, are different in direction but exactly symmetrical with those from A through D that we traced above. As was shown for the movement from A to the intersection, the continuous path that the economy might follow from E to H and then to the intersection is indicated by a dashed line. Once at the intersection, we have the combination of Y and r that provides equilibrium in both markets; the income level and the interest rate will remain unchanged until the existing equilibrium is upset by a shift in either the IS or LM curve or both.[8]

CHANGES IN THE TWO-MARKET EQUILIBRIUM

The equilibrium combination of Y and r identified by the intersection of the IS and LM functions will, of course, change in response to any shift in those functions. Shifts in the IS function are caused by shifts in the investment function or the saving function (Parts A and C of Figure 16-3); shifts in the LM curve are caused by shifts in the money supply or transactions demand or speculative demand functions (Parts B, C, and A, respectively, of Figure 16-4). Finally, a shift in any of the functions on which the IS and LM curves are based may result from a change in any of the factors that determine the positions of these functions. We thus have a method of analysis by which we can trace the effects of a change in any of these many underlying factors through the system to its final effect on the income level and interest rate—assuming, of course, that all other factors remain unchanged.

[8]Although it is quite illuminating to trace the path followed by Y and r in the way we have done here, it must be recognized that the IS–LM model now before us does not in itself tell us that Y and r will follow the path here described or any other particular path from an initial disequilibrium position like A or E. As was briefly explained in Chapter 3, to trace the process of change in the values of a model's variables from one period to another can be done only with a dynamic model—the IS–LM model is not dynamic but completely static. It can and does identify the values the variables must exhibit in order that there be equilibrium, but it does not show the sequence of changes by which these values will be reached if we start off with any values other than these equilibrium values.

A Change in Investment

Among the various possibilities, a shift in the investment or MEI curve is one of the most important. Suppose a change in an underlying factor, e.g., an improvement in business expectations, causes this curve to shift \$20 to the right *at each rate of interest*. In Part A of Figure 16-7, the original curve is labeled MEI_1 and the new one MEI_2. Just as an *IS* curve was derived graphically in Figure 16-3 from the investment and saving curves given there, so in Figure 16-7 one may derive a separate *IS* curve from each of the investment curves in combination with the given saving curve. In Part D, the IS_1 and IS_2 curves are based on the MEI_1 and MEI_2 curves, respectively. Examination shows that IS_2 lies \$40 to the right of IS_1 *at each interest rate*. In other words, at each interest rate the level of income at which $S = I$ is now \$40 greater than it was before the shift in the investment schedule. This follows from the fact that, with an increase of \$20 in investment, income must rise by \$40 to induce an increase of \$20 in saving, given that the MPS is 1/2. This is nothing more than the simple multiplier, $\Delta Y = \Delta I/\text{MPS}$, which in the present case gives us $\$40 = \$20/0.5$.

The original equilibrium was earlier found at *Y* of \$120 and *r* of 3 percent and is shown here again by the intersection of IS_1 and *LM* in Part D of Figure 16-7. This gives us the AD_1 curve positioned at *Y* of \$120 in Part E. The *LM* curve here is the same as the one derived in Figure 16-4. The new equilibrium that results from the shift in the investment curve is at *Y* of \$140 and *r* of 3.5 percent as shown by the intersection of IS_2 and *LM* in Part D. If the interest rate had not risen at all, income would have risen by \$40 or from \$120 to \$160. But we do not get this rise in income. Instead, as the increase in investment spending starts an upward movement in income, the rising income level increases the money balances needed for transactions purposes. This leads to a rising interest rate, which in turn feeds back to make the increase in investment spending less than the \$20 and the increase in income less than the \$40 they would have been with no rise in the interest rate. What we find in the present illustration is that, with the *LM* curve as given, the shift in the *IS* curve caused by a rightward shift of \$20 in the investment curve raises *Y* by \$20 and *r* by 0.5 percentage point. The \$20 rise in *Y* means an increase of \$10 in required transactions balances, given $k = 0.5$, and the rise in *r* of 0.5 percentage point is just the rise needed to reduce the amount of money the public wishes to hold in speculative balances by \$10, thus supplying the additional \$10 needed for transactions balances. With $\Delta Y = \$20$ and $\Delta r = 0.5$ percentage point, therefore, the supply of and demand for money will again be in balance.

The same 0.5 percentage point rise in *r* will cut the *increase* in investment from \$20 to \$10. As may be seen in Part A of Figure 16-7, investment, which would have risen from its original \$20 at *r* of 3 percent (point *E*) to \$40 with no change in *r* (point *F*), instead only rises from \$20 to \$30 (point *G*) as one-half of what would have been the larger increase of \$20 is choked off by the rise in *r* from 3 to 3.5 percent. The final increase in investment of \$10 turns out to be the same amount as the increase in saving that occurs with a rise in income of \$20. From $\Delta S = s(\Delta Y)$, we here have $\$10 = 0.5(\$20)$. With $\Delta Y = \$20$ and $\Delta r = 0.5$ percentage point, *S* will again equal *I* and *Y* will again equal $C + I$. No other combination of changes in *Y* and *r* but this particular one will be consistent with equilibrium in both the goods and money markets, assuming the indicated rightward shift in the investment schedule with all other things remaining as given.

For a leftward shift in the investment schedule, the results will, of course, be the opposite. In the present illustration, if the investment schedule of Part A were to shift to the left by \$20 or from MEI_1 to MEI_3, the new equilibrium would be found in Part D with *Y* of \$100 and

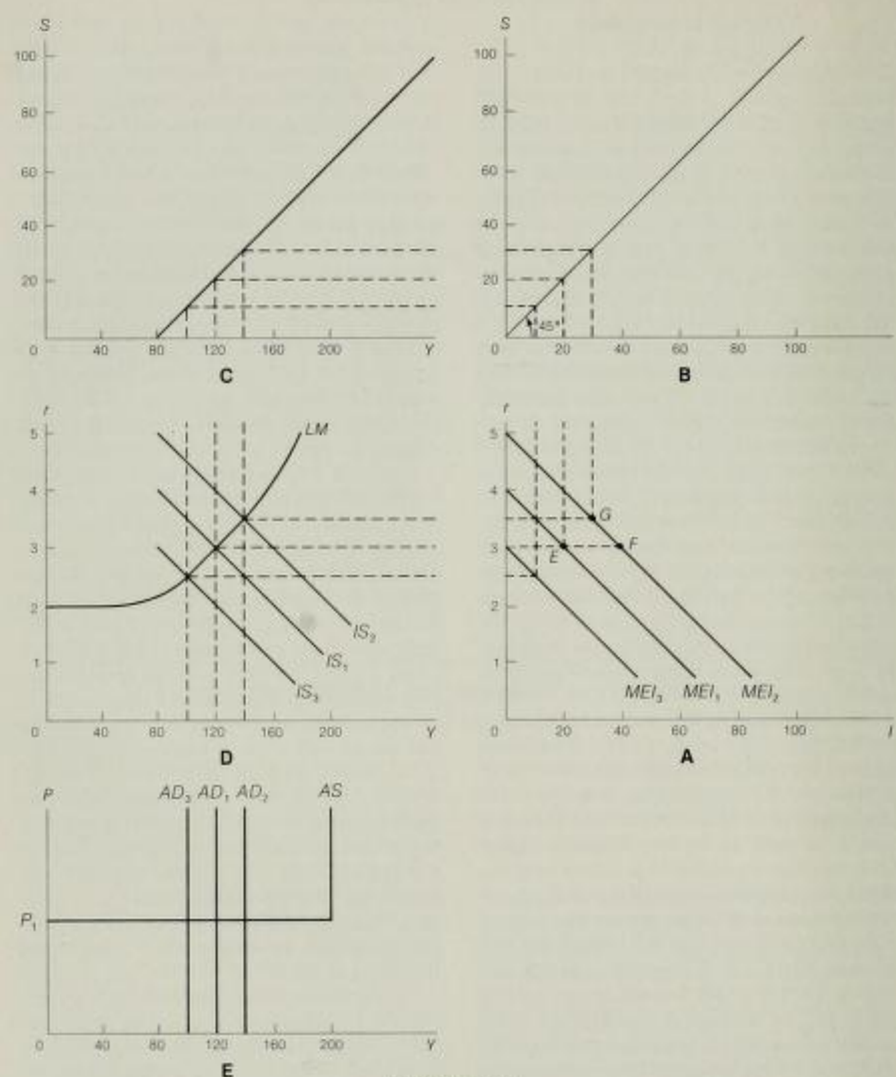


FIGURE 16-7
A change in investment and the change in aggregate demand

r of 2.5 percent or, relative to the original equilibrium, a decrease in Y of \$20 and in r of 0.5 percentage point.[9] By the same reasoning used above, it may be seen that no other combination of changes in Y or r but the one noted would be consistent with equilibrium in both the goods and money markets, assuming the indicated shift in the investment schedule with all other things remaining as given.

Just as the intersection of IS_1 and LM in Part D established the AD_1 curve in Part E, the intersections of the IS_2 and LM curves and of the IS_3 and LM curves establish the AD_2 and AD_3 curves, respectively, in Part E. Given the LM curve in Part D, the shifts in the MEI curve of Part A which cause the indicated shifts in the IS curve in Part D also produce the shifts here noted in the AD curve in Part E.

A Change in the Money Supply

As a second illustration, let us take the case of a \$20 increase in the money supply. This shifts the m_s curve in Part B of Figure 16-8 from its original position m_{s_1} to the new position m_{s_2}. With no change in the speculative demand function of Part A and no change in the transactions demand function of Part C, it is found that the \$20 increase in m_s shifts the LM function rightward by \$40 *at each rate of interest*, or from LM_1 to LM_2. What lies behind this may be seen as follows. Equilibrium between m_d and m_s requires a rise in Y sufficient to absorb the \$20 increase in m_s in transactions balances, m_t, if the interest rate is assumed to be given. Since $m_t = k(Y)$, we have $Y = m_t/k$ and $\Delta Y = \Delta m_t/k$. Accordingly, with k given as 0.5, ΔY must be \$40 to produce a new equilibrium between m_d and m_s at each interest rate.

The original equilibrium was found earlier at Y of \$120 and r of 3 percent and is shown here again by the intersection of IS and LM_1 in Part D of Figure 16-8. The IS curve here is the same as the one originally derived in Figure 16-3. The new equilibrium that results from the increase in the money supply is at Y of \$140 and r of 2.5 percent. Although the \$20 increase in m_s will shift the LM curve \$40 to the right at each interest rate, it will not raise the equilibrium level of Y by \$40, because, with no shift in the IS curve, a rise in the equilibrium level of income cannot occur unless there is a fall in r. However, a fall in r will increase the amount of money people choose to hold in speculative balances. In the present illustration, it turns out that \$10 of the \$20 increase in M will be absorbed in speculative balances as r falls from 3 to 2.5 percent (as may be seen from Part A of Figure 16-8). This same fall in r is also just the fall that will raise I by \$10 (as may be seen from Part A of Figure 16-7) and, through the multiplier, raise Y by \$20. A rise in Y of \$20 increases required transactions balances by \$10, which accounts for the balance of the \$20 increase in m_s. No other possible combination of changes in Y and r but this +\$20 and −0.5 percentage point will be consistent with equilibrium, assuming the indicated increase in the money supply with all other things remaining equal.

If we took the opposite change, a \$20 decrease in the money supply that shifts the m_s curve from m_{s_1} to m_{s_3}, we would find an opposite result. The new equilibrium combination of Y and r would be at \$100 and 3.5 percent, or a decrease in Y of \$20 and a rise in r of 0.5 percentage point. By applying the reasoning of the preceding paragraph, one may see that no other combination of changes in Y and r but this particular one will provide a new equilibrium within the assumptions of the present illustration.

As the LM_1 and IS curves in Figure 16-8 are identical with the LM and IS_1 curves in Figure

[9] For simplicity it is here being assumed that the LM curve is linear over the range of interest rates (2.5 to 3.5 percent) relevant to our illustrations. (This requires that we also assume that the underlying speculative demand curve is linear over this same range.) Specifically, the slope of LM is taken to be 0.025 (r changes by 0.025 percentage point for each \$1 billion change in Y or by 0.5 percentage point for a \$20 billion change in Y). Any departure from linearity would give actual numerical results different from the symmetrical ones found in the illustrations.

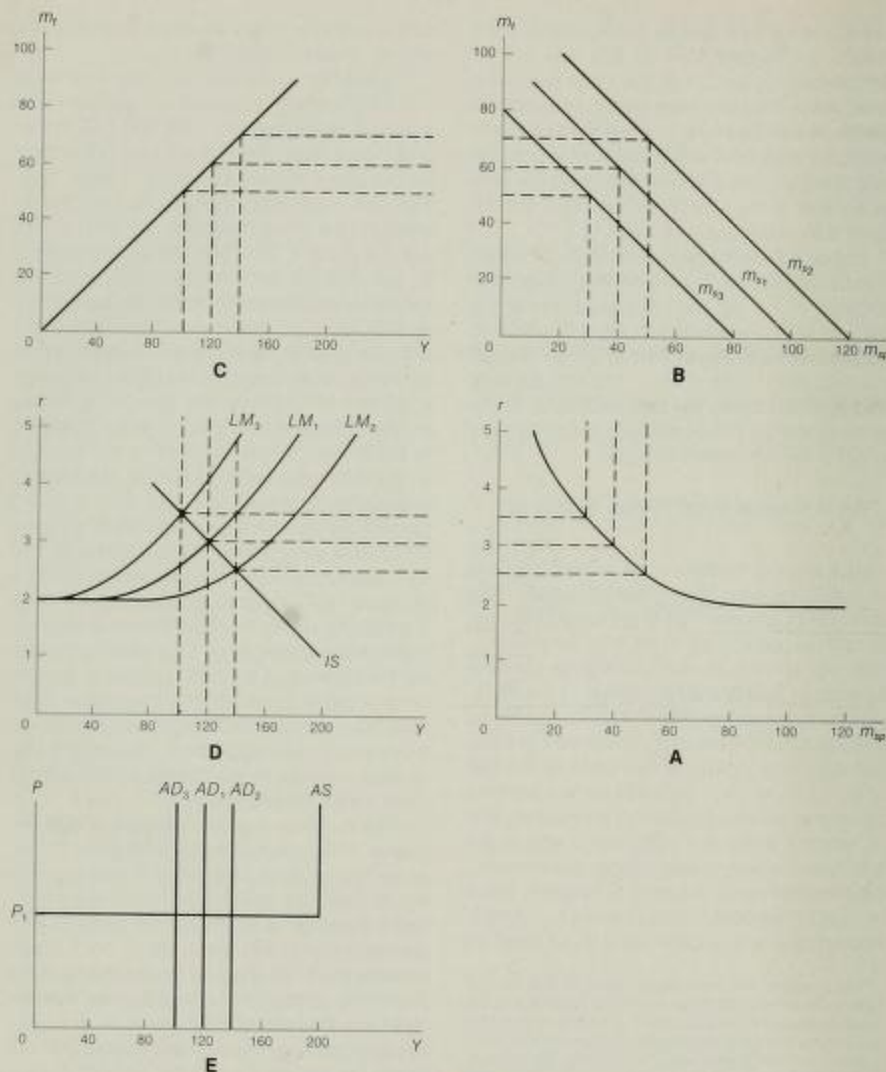


FIGURE 16-8
A change in the money supply and the change in aggregate demand

from \$120 to \$160, with an increase in investment of \$20 and an MPC of 1/2, is just the result we found in the simple Keynesian model in Chapter 5. Now we see that this result will be realized only if an appropriately expansionary monetary policy, here an increase in m_s of \$20, is pursued to prevent what otherwise would be a rise in the interest rate and so a smaller rise in the income level.

The effects of shifts in other functions may be traced in the same way. For example, an increase in "thrift," which appears as an upward shift in the saving function (Part C of Figure 16-3), will shift the *IS* curve to the left and lower *r* and *Y*. An increase in the demand for money to be held in idle balances, which appears as a shift to the right in the speculative demand function (Part A of Figure 16-4), will shift the *LM* curve to the left, raise *r*, and lower *Y*. A change in payments practices that makes it possible for each dollar of money to handle a larger volume of transactions per time period reduces *k* and appears as a less steeply inclined transactions demand function (Part C of Figure 16-4). This will shift the *LM* curve to the right, lower *r*, and raise *Y*.

GOVERNMENT SPENDING, TAXATION, AND TWO-MARKET EQUILIBRIUM

In now adding government spending and taxation to the general equilibrium model, the equilibrium condition in the goods market for a two-sector economy, $S = I$, becomes $S + T = I + G$ for a three-sector economy. This simply means that the aggregate spending for goods and aggregate output of goods will be equal when the sum of the diversions, $S + T$, from the real income stream is just matched by the sum of compensating injections, $I + G$, into the real income stream. Alternatively the equilibrium condition in the goods market may be expressed as $Y = C + I + G$. The equilibrium condition in the money market is $m_d = m_s$ as before.

As in the first fiscal model of Chapter 6, government purchases of goods and services and net tax receipts are assumed to be independent of the level of income. Part A of Figure 16-10 shows \$20 of government purchases added to the investment schedule of Figure 16-3. Since these purchases are regarded as also independent of the interest rate, the $I + G$ curve lies \$20 to the right of the *I* curve at all interest rates. Whatever the interest rate, the sum of *I* and *G* will be \$20 greater than *I* alone. In terms of its effect on *Y*, a dollar of *G* is no different from a dollar of *I*. Adding \$20 of *G* thus shifts the *IS* curve \$40 to the right, from IS_1 to IS_2, for the same reason that the increase in investment of \$20 discussed earlier shifted the *IS* curve to the right by \$40 in Figure 16-7.[10] Part D of Figure 16-10 includes the same *LM* function derived in Figure 16-4.

Other things being equal, the introduction of deficit-financed government purchases of \$20 moves the *Y*, *r* equilibrium in Part D from \$120 and 3 percent to \$140 and 3.5 percent and shifts the *AD* curve in Part E from AD_1 to AD_2. Again the result shown is the same as that in Figure 16-7 for a \$20 shift in the investment demand schedule. What otherwise would be an expansion in *Y* of \$40, as indicated by the simple multiplier of 2, becomes the lesser expansion of \$20 due to the effect of the rise in *r* that accompanies the rise in *Y*. However, there is this difference. *G* of \$20 is unaffected by the

[10] It would be more correct to designate the curve as *IG-ST* instead of *IS*, but the simpler notation will be retained. Note, however, that in Parts A–C the axes previously labeled *I* are now $I + G$, and the axes previously labeled *S* are now $S + T$.

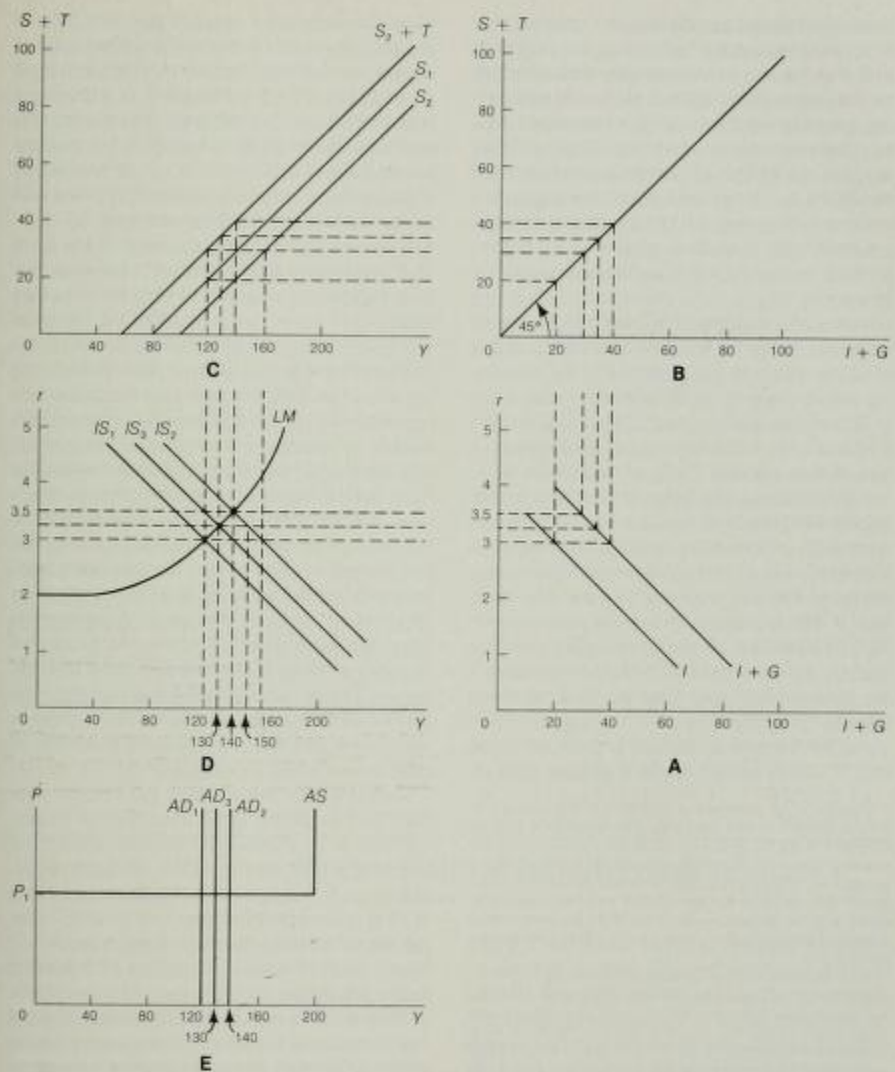


FIGURE 16-10
Effect on aggregate demand of changes in government spending and taxation

rise in r it causes, but the rise in r reduces I by \$10, which makes the net change in $I + G$ only \$10 and the rise in income only \$20. The full income-expansionary effect of G is not realized, because the resulting rise in r chokes off \$10 of private investment spending. Thus, *a fiscal policy that is designed to raise the income level through a deficit-financed expansion of government spending may not produce the maximum possible rise of income unless it is accompanied by an appropriately expansionary monetary policy.*[11]

Let us now suppose that there is a balanced budget and that the government collects taxes of \$20 to match its spending of \$20, thus avoiding deficit spending. In the present model, taxes of \$20 reduce disposable income by \$20. With the MPS of 1/2, the reduction in saving is one-half of this amount. Thus, at each level of Y, T of \$20 reduces S by \$10 and C by \$10, which appears in Part C of Figure 16-10 as a downward shift of \$10 in the saving function, or a shift from S_1 to S_2. To the leakage from income made up of saving must now be added the leakage of \$20 for taxes. This gives us the curve $S_2 + T$, the sum of saving and taxes, or that portion of the income flow that does not appear as consumption spending at each level of income.[12]

I of Part A and S_1 of Part C gave us IS_1 of Part D; $I + G$ of Part A with T of zero gave us IS_2 of Part D; finally, $I + G$ of Part A with $S_2 + T$ of Part C gives us IS_3 of Part D. The new equilibrium position indicated by the intersection of IS_3 and LM in Part D is found at Y of \$130 and r of 3.25 percent. Corresponding to this is AD_3 positioned at Y of 130 in Part E. In our illustration, adding G of \$20 and an equal amount of T raises the equilibrium level of Y by one-half the increase in the size of the budget.[13]

With G and T both independent of the level of Y, we have a model similar to the one that gave us the unit multiplier in Chapter 6. In that model, the rise in Y was equal to the increase in the size of the budget. However, since the interest rate is now part of the model, we find that the actual multiplier is less than the balanced-budget multiplier of 1 that appears in the simpler model. An expansion in the size of the budget, with the budget balanced, will raise the income level, but the rise in income, which would otherwise be equal to the expansion in the size of the budget, will be dampened by the tendency for the interest rate to rise with the rise in income. In other words, *a fiscal policy designed to produce a rise in income while maintaining a balanced budget will produce the maximum possible income increase only if it is accompanied by an expansionary monetary policy that prevents what otherwise might be a rise in the interest rate and a consequent reduction in private investment spending.*

We have seen that a rise in the income level may be expected from an expansion in G with no change in T and even from an expansion in G that is matched by T. The third possibility, of course, is a reduction in T with no reduction in G, a commonly cited example of which was the tax cut of 1964. This was a major reduction which cut federal tax receipts about 10 percent below what they otherwise would have been; in contrast, the more recent anti-recessionary Tax Reduction Act of 1975 reduced receipts about 5 percent below what they otherwise

[11]Because the government spending in this example is entirely deficit financed, we are concerned with the method of deficit financing employed. If entirely financed by the sale of government securities to the public, there will be no increase in the money supply; the results are as described above. If financed by the appropriate "mix" of sales to the public and the banking system, there will be an increase in the money supply that permits the full \$40 potential expansion in Y.

[12]For example, with Y of \$140 and T of zero, Y_d, or $Y - T$, would be \$140; C would be \$110, or \$40 + ½(\$140 − 0); and S would be \$30, or −\$40 + ½(\$140 − 0), the last figure as shown on the S_1 curve of Part C of Figure 16-10 at Y of \$140. The imposition of T of \$20 reduces Y_d to \$120 when Y is \$140. This reduces C to \$100, or \$40 + ½(\$140 − \$20), and S to \$20, or −\$40 + ½(\$140 − \$20), the latter figure as shown on the S_2 curve at Y of \$140. Finally, adding T of \$20 makes total diversions from income \$40 at Y of \$140, as shown on the $S_2 + T$ curve.

[13]The original budget was one in which both G and T were zero.

would have been. Figure 16-10 may be used to illustrate an aspect of the 1964 tax cut much discussed at the time. Suppose the original equilibrium is that defined by the intersection of IS_3 and LM at Y of \$130 and r of 3.25 percent; this is the equilibrium consistent with $I + G$ of Part A and $S_2 + T$ of Part C of Figure 16-10. With no change in G but a tax cut of \$20, the $I + G$ curve remains as is, and the $S_2 + T$ curve shifts downward to S_1. This in turn causes the IS curve to shift from IS_3 to IS_2. But the full expansionary effect of the tax cut—a rise in Y from \$130 to \$150—is not realized because the interest rate rises. Hence, in judging the prospective effectiveness of the 1964 tax cut, one consideration was whether the increase in aggregate spending that would follow therefrom would be smaller than otherwise obtainable due to adverse monetary effects. In the late President Johnson's words, "It would be self-defeating to cancel the stimulus of tax reduction by tightening money. Monetary and debt policy should be directed toward maintaining interest rates and credit conditions that encourage private investment."[14] The model in Figure 16-10 is far too simple to come to grips with the questions involved, but it suggests, in very general terms, that what is called for is an increase in the money supply sufficient to shift the LM curve to the right by the amount necessary to secure the greater rise in income—from \$130 to \$150—that will follow from the increase in aggregate spending to be expected at a stable interest rate.

Although we will not go beyond the simple model in which both G and T are assumed to be independent of Y, it should be noted that the IS–LM analysis of Figure 16-10 may be elaborated by introducing more realistic fiscal assumptions. In Part C, for example, T may be treated as a function of Y, and the effects of this more realistic fiscal assumption on the Y, r equilibrium combination may readily be traced. This model will show how the potential income-expansionary effect of, say, a rise in investment spending may be restrained by both a rise in the interest rate and a rise in tax receipts as income expands. Though it adds something to the simpler model of this section, like any other model of this kind it will again bring out the principal conclusion emphasized here: An increase in aggregate spending, whether the result of a shift in the investment function or consumption function or of a change in government spending or taxation, will not produce the effect on income suggested by the crude multipliers in earlier chapters. When we recognize the role played by money and interest, we see how an otherwise greater expansion of income suggested by crude multipliers may be prevented by the rise in the interest rate that may accompany a rise in income.

THE IS AND LM ELASTICITIES AND MONETARY-FISCAL POLICIES

So far we have intentionally avoided specific reference to the elasticities of the IS and LM functions so that we might concentrate on the general characteristics of the present stable-price model and the general conclusions it suggests. In now taking account of the elasticities of these functions, we will find that some of these conclusions must be qualified and that some must even be abandoned in the extreme cases of perfectly elastic or inelastic functions. For example, it is possible, as we shall see in a moment, that an expansionary fiscal policy will raise only the interest rate and leave the income level unchanged or that it will raise only the income level and leave the interest rate unchanged. It is possible that an expansionary

[14] *Economic Report of the President*, Jan. 1964, p. 11.

monetary policy will lower only the interest rate and leave the income level unchanged or that it will change neither the interest rate nor the level of income. The reverse is possible for contractionary policies.

Elasticity of the IS and LM Functions

With a fixed money supply, the *LM* function as derived in Figure 16-4 slopes upward to the right. However, at one extreme the function may be expected to become perfectly elastic, and at the other extreme it may be expected to become perfectly inelastic, with a range of varying elasticities in between. In general, the higher the interest rate, the less elastic the corresponding point on the *LM* function will be. These three ranges are laid off in Part A of Figure 16-11, in which the perfectly elastic section is the "Keynesian range," the perfectly inelastic section is the "classical range," and the section between is the "intermediate range."

Why this particular shape with perfect elasticity at one extreme and perfect inelasticity at the other? Recall that at some very low interest rate the speculative demand for money may become perfectly elastic, the result of a consensus by wealth-holders that the interest rate will fall no lower and that security prices will rise no higher. Wealth-holders accordingly stand ready to exchange securities for cash at existing security prices, which produces the liquidity trap on the speculative demand function. Here, on the *LM* function, it produces what is known as the Keynesian range. At the other extreme, at some very high interest rate, the speculative demand for money may become zero and perfectly inelastic at interest rates above this, the result of a consensus by wealth-holders that the interest rate will rise no higher and that security prices will fall no lower. At this or any higher rate, wealth-holders accordingly prefer to hold only securities and no idle cash. This perfectly inelastic section of the speculative demand function is what is known as the classical range on the *LM* function.

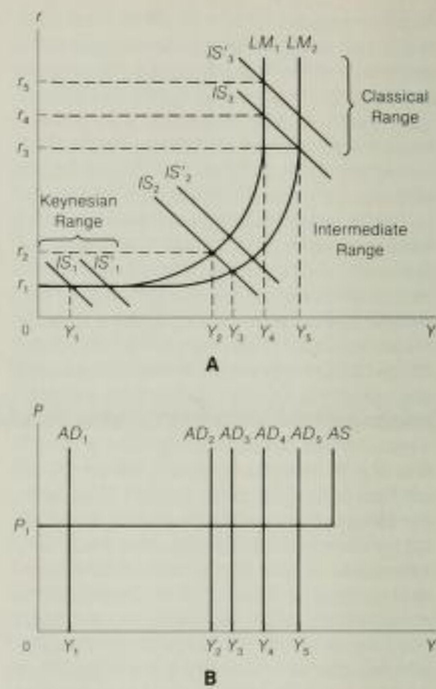


FIGURE 16-11
Effects of shifts in the IS and LM functions with various elasticities of the LM function

Why are the three sections into which the *LM* function has been divided labeled in this fashion? Recall that, in our simplified version of the classical theory, money is demanded only for transactions purposes. Thus, in Figure 16-4, classical theory assumes that the speculative demand for money is zero at each interest rate. In effect, Part A of that figure vanishes. If the total money supply given in Part B is \$100, that \$100 will be held in transactions balances or $m_{sp} = 0$ and $m_s = m_t$. With k given in Part C as 1/2, the *LM* curve of Part D becomes a perfectly vertical line at the income level of \$200. If the public holds money only for transactions

purposes and if it holds money balances equal to one-half of a period's income, money market equilibrium is found at an income level of \$200 at all interest rates.[15]

With the exception of the perfectly inelastic section, or the so-called classical range, it would not be altogether incorrect to include the remainder of the *LM* function in the Keynesian range. However, because of Keynes's emphasis on the ineffectiveness of monetary policy, the liquidity-trap section alone has come to be identified as the Keynesian range. Within this range, monetary policy is completely ineffective; therefore, this range most closely fits Keynes's emphasis.

The *IS* function as derived in Figure 16-3 slopes downward to the right. Its elasticity depends on the responsiveness of investment spending to changes in the interest rate and on the magnitude of the multiplier. If the investment spending schedule is perfectly inelastic, indicating that investment spending is completely insensitive to the interest rate, the *IS* curve derived in Part D will be perfectly inelastic, regardless of the magnitude of the multiplier. If, on the other hand, the investment demand schedule shows some elasticity, as seems to be the case, the *IS* curve will be more elastic, the lower the MPS. The lower the MPS, the higher will be the multiplier and so the greater will be the change in income for any increase in investment resulting from a fall in the interest rate. Part A of Figure 16-12 shows three pairs of *IS* curves, each made up of one highly inelastic and one elastic *IS* curve.

Parts B of both Figures 16-11 and 16-12 show *AD* curves corresponding to the various levels of *Y* set off on the *Y* axis in Part A of each of those figures. As in previous figures of this chapter, the *AS* curve is assumed to be perfectly elastic up to the full-employment level of output. Also as before, the full-employment level is assumed to be greater than the highest level of output attained, Y_5 in Part A, so that all of the changes in *AD* shown in each figure from Y_1 to Y_5 are accompanied by proportional changes in *Y*.[16]

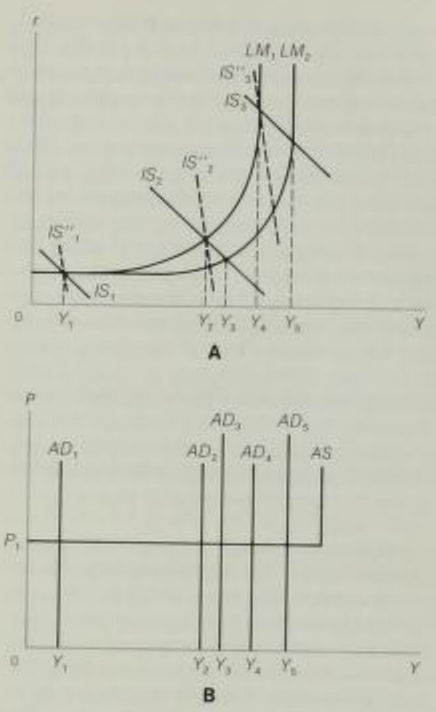


FIGURE 16-12
Effects of elastic and inelastic IS functions in different ranges of the LM function

[15]The graphic derivation of a perfectly inelastic *LM* curve is shown in Chapter 17 on p. 316.

[16]Although the present model contains a classical element in the form of the perfectly inelastic range of the *LM* curve, the model is essentially Keynesian as it shows that the equilibrium level of output may be below the level consistent with full employment. Recall from Chapter 14 that the simple classical model with its assumption of perfect wage and price flexibility yields a perfectly inelastic *AS* curve, which is located at the full-employment level of output. This makes the full-employment level of output the only equilibrium level, a result altogether different from that found in Figures 16-11 and 16-12. Further comparisons between the classical and Keynesian models in terms of the *IS-LM* framework will be presented in the following chapter.

Monetary and Fiscal Policy

Monetary policy is the exercise of the central bank's control over the money supply as an instrument for achieving the objectives of general economic policy. Fiscal policy is the exercise of the government's control over public spending and tax collections as instruments for the same purpose. We will confine ourselves here to the single policy objective of raising the level of real income. The *IS–LM* framework then provides a basis for comparing the effect of the two types of policy on the income level and the interest rate and for comparing conditions under which each type of policy will be effective or ineffective in producing the change in income that is the policy objective. For this purpose, the discussion is conveniently divided into three parts, each corresponding to a range of the *LM* function in Part A of Figure 16-11.

The Keynesian Range Consider first the Y_1, r_1 equilibrium in the Keynesian range. An increase in the money supply shifts the *LM* curve to the right, from LM_1 to LM_2, and means that *for each possible level of income* $m_s = m_d$ only at a lower interest rate; the rate must fall by the amount necessary to make the public willing to hold larger idle cash balances. But this is not true in the "liquidity trap." Here the interest rate is already at what is for the time being an irreducible minimum. As the monetary authority purchases securities, security-holders are willing to exchange them for cash at the existing prices of securities. Therefore, expansion of the money supply cannot cause the interest rate to fall below the rate given by the trap. All that happens is that the public holds more in speculative balances and less in securities. Further increases in the money supply would be expected to shift the *LM* curve still farther to the right, but the lower end of the curve will remain anchored in the same liquidity trap. If the economy is already in the trap, it follows that monetary policy is powerless to raise the income level, since it cannot reduce the interest rate any further and thereby produce a movement down the IS_1 curve to a higher equilibrium income level. The belief that the economy was in the trap during the early thirties led Keynes to his then unorthodox fiscal policy prescriptions. Since government cannot raise the income level through monetary policy, whatever government is to do through monetary-fiscal policy it must then do through fiscal policy. If a rise in income cannot be achieved by producing a movement down the IS_1 curve through monetary expansion, it can be achieved by producing a shift in the IS_1 curve itself, say from IS_1 to IS'_1. Fiscal measures such as increased government spending or reduced taxes that could shift the *IS* curve became the order of the day.

To the extent that monetary policy operates by raising investment spending through a reduction in the cost of money, the impasse of monetary policy for an economy caught in the trap means that the elasticity or inelasticity of the *IS* function is no longer relevant. In Part A of Figure 16-12, for example, it does not matter whether the *IS* function is the elastic IS_1 or the inelastic IS''_1.[17]

The liquidity trap is an extreme case that could occur only during a deep depression, if even then. A prosperous economy and a liquidity trap do not go hand in hand. Since the pure Keynesian range is the range of the liquidity trap, one can now appreciate what Professor Hicks meant by his observation, made shortly after the appearance of Keynes's book, that "the General Theory of Employment is the Economics of Depression."[18]

The Classical Range Next let us examine the Y_4, r_4 equilibrium defined by the intersection of IS_3 and LM_1 in Part A of Figure 16-11. There is some increase in the money supply that will shift the LM_1 curve to LM_2. In contrast to the result in the Keynesian range, the result is now an increase in the income level from Y_4 to Y_5 and a fall in the interest rate from r_4 to r_3. In the

[17] As we will see, the elasticity of the *IS* function does become relevant elsewhere, but not in the Keynesian range.
[18] J. R. Hicks, *op. cit.*, p. 472.

classical range, the interest rate is so high that speculative balances are zero; money is held for transactions purposes only. If the monetary authority under these circumstances enters the market to purchase securities, security-holders can be induced to exchange securities for cash only at higher security prices. As security prices are bid up and the interest rate is pushed down, investment is stimulated (and, in classical theory, saving is discouraged). Since nobody chooses to hold idle cash, expansion of the money supply will produce a new equilibrium only by reducing the interest rate by whatever amount is necessary to increase the income level sufficiently to absorb the full increase in the money supply in transactions balances. If in the present case we assume that $\Delta m_s = \$20$ and $k = 1/2$, equilibrium will be restored only when Y has risen by \$40, or, in general, when $\Delta Y = \Delta m_s/k$. In the classical range, the result follows the simple classical quantity theory of money as a theory of aggregate demand. Y rises proportionally with the increase in m_s. If $V = 2$ or $k = 1/2$, the rise in Y must be twice the rise in m_s in order to satisfy the equilibrium condition: $m_sV = Y$ and $m_s = k(Y)$.

In contrast to the Keynesian range, in which monetary policy is completely ineffective, in the classical range it appears to be completely effective. No part of any increase in the money supply disappears into idle cash balances. The increase in the money supply means increased spending that produces a rise to that income level at which the total increase in the money supply is absorbed into transactions balances. Because all income changes are real changes in the present model, we find that the increase in the money supply that shifts LM_1 to LM_2 causes an increase from Y_4 to Y_5 in output as well as in income.

Again in contrast to the Keynesian range, in which fiscal policy alone can be effective, fiscal policy in the classical range is completely ineffective. An upward shift in the *IS* function from IS_3 to IS'_3 in Part A of Figure 16-11 can raise only the interest rate, from r_4 to r_5; the income level stays unchanged at Y_4. Given the increase in spending that lies behind the upward shift in the *IS* function, there will be a rise in the rate of interest sufficient to choke off enough spending to leave aggregate spending unchanged. Thus, if the rise in spending resulted from increased government spending, the rise in the interest rate would choke off an amount of private spending equal to the rise in government spending. The level of income is as high as the given money supply can support. In the classical range, an increase in income is thus impossible without an increase in the money supply, and monetary policy becomes an all-powerful method of controlling the income level.

How does the elasticity of the *IS* function affect the equilibrium positions in the classical range? Let us compare the elastic IS_3 function and the inelastic IS''_3 function shown in Part A of Figure 16-12. Here we see that with the IS''_3 function no increase in the money supply and no reduction in the interest rate is capable of raising the income level from Y_4 to Y_5. Monetary policy will raise Y but not by the multiple of m_s given by $1/k$. Although this seems to upset the result suggested by classical theory, classical theorists would deny that the *IS* curve could be so inelastic. Recall that in both classical and Keynesian theory investment is a function of the interest rate, but in classical theory saving also is a function of the interest rate. Thus, it can be shown that only if both saving and investment are quite insensitive to the interest rate could there be an inelastic curve of the sort described by IS''_3 in Part A of Figure 16-12.[19] As long as one or the other is elastic, the resulting *IS* function will also be elastic, and with an elastic *IS* function the result of a change in the money supply in the present model is $\Delta Y = \Delta m_s/k$.

[19] In terms of Part C of Figure 16-3, we may show saving as a function of both Y and r by drawing in a similar fashion sucessively higher saving functions to correspond with successively higher interest rates. An inelastic investment function in Part A combined with this income-elastic and interest-elastic saving function in Part C will still produce an elastic *IS* function in Part D.

The Intermediate Range Finally, let us examine the equilibrium of Y_2, r_2, as defined by the intersection of IS_2 and LM_1 in Part A of Figure 16-11. Here again we see that there is some increase in the money supply that will shift the LM_1 function to LM_2. In the Keynesian range, this increase in the money supply left both Y and r unchanged because that total increase was absorbed in speculative balances at the existing interest rate, which defines the liquidity trap. In the classical range, this increase in the money supply raised Y by the amount necessary to absorb the full increase in transactions balances. This worked itself out through that reduction in the interest rate that raised spending by the amount needed to produce the required rise in income. In the intermediate range, however, the increase in the money supply is absorbed partially in speculative balances and partially in transactions balances. The level of income rises but by an amount less than that which would require the full increase in the money supply for transactions purposes.

To illustrate, let us suppose that the increase in the money supply is \$20 and k is 1/2. The resultant shift in the LM function is \$40, but in this case the rise in income $(Y_3 - Y_2)$ is only half that amount. In reducing the interest rate by the amount that produces the increase in spending needed to raise the income level by \$20, \$10 (one-half of the increase in the money supply) is absorbed in speculative balances. The remaining \$10 is just the additional amount of money needed for transactions purposes with the income level up by \$20.

Thus, in the intermediate range, monetary policy is found to have a degree of effectiveness but not the complete effectiveness it has in the classical range. In general, the closer the equilibrium intersection is to the classical range, the more effective monetary policy becomes, and the closer the intersection is to the Keynesian range, the less effective it becomes.

Within this range, fiscal policy is also effective to some extent. Fiscal measures that shift the IS function from IS_2 to IS'_2, for example, will raise the level of income and the interest rate to the new equilibrium defined by the intersection of IS'_2 and LM_1. If the shift in the IS function is the result of a deficit-financed increase in government spending, the interest rate must rise. We are assuming a fixed money supply described by LM_1, so the increased government spending is being financed by borrowing from the public. The sale of additional securities by the government depresses security prices, raises the interest rate, and chokes off some amount of private spending. The rise in the interest rate following any given increase in government spending will be greater the higher in the intermediate range the equilibrium happens to be. Conversely, it will be smaller the lower in the intermediate range the equilibrium happens to be. Although fiscal policy is found to have a degree of effectiveness anywhere in the intermediate range, in general it will be more effective the closer equilibrium is to the Keynesian range and less effective the closer equilibrium is to the classical range.

Although both monetary and fiscal policies have varying degrees of effectiveness in the intermediate range, the relative effectiveness of each depends in large part on the elasticity of the IS function. If the IS function is the inelastic IS''_2 in Part A of Figure 16-12, monetary policy can do very little to raise the level of income, even in the intermediate range; fiscal policy alone is effective in such a situation. Furthermore, an expansionary fiscal policy need not be concerned with adverse monetary effects in this case. A shift in an inelastic IS function will raise the interest rate, but that will have little feedback on spending. Keynes maintained that the investment schedule (as well as the saving schedule) was interest inelastic. If this is the case, the IS schedule must also be inelastic, and fiscal policy, which is completely effective in the Keynesian range, would be almost as effective in the intermediate range. If the IS schedule is indeed interest inelastic, then the Keynesian range becomes, in effect, the complete LM curve, more applicable at the lower end than at the upper end but with some applicability throughout.

chapter

17 The Extended Model: Variable Price Level

We saw at the end of the preceding chapter that either fiscal or monetary policy may be employed in varying circumstances to shift the *IS* and *LM* functions to the right and thereby to shift the *AD* function to the right and raise the equilibrium level of *Y*. Because our model assumed that the *AS* curve is perfectly elastic up to the full-employment level of income, each of the illustrated policy-generated shifts in the *AD* curve along that *AS* curve identified a new equilibrium level of *Y* at an unchanged *P*.

The assumption of this kind of *AS* function must be recognized as no more than a convenient simplification, although in the early days of Keynesian economics it was accepted as much more than that. There was a belief then that changes in *AD* would be without an effect on *P* over a range of *Y* from a depression level to a level practically up to the full-employment level. There was not then and there is not today much doubt that an increase in *AD* will leave *P* unaffected if that increase occurs under depression conditions of the severity experienced during the early nineteen-thirties. The *AS* curve is perfectly elastic or virtually so with the economy operating so far to the left along that curve. However, there is little doubt today that an increase in *AD* will have some upward effect on *P*, even though that increase still leaves the level of output somewhat below the full-employment level. The difficulty is in distinguishing the point at which the curve begins to slope upward and how fast it slopes upward. For example, in late 1976 and early 1977, with the unemployment rate near 8 percent and the rate of capacity utilization about 75 percent, some economists asserted that fiscal and monetary actions to shift the *AD* curve rightward would result almost entirely in lower unemployment and greater output and in only negligible, if any, pressure on prices; other economists who emphasize considerations other than the spread between actual and potential aggregate output saw the makings of a sharp rise in prices as a result of the very same actions.

For present purposes, it is only necessary to recognize that the *AS* curve does display an upward-sloping range between a perfectly elastic range at depression levels of output and a perfectly inelastic range at the full-employment level of output. As was seen in the derivation of the different kinds of *AS* curves in Chapter 13, the upward slope must necessarily be a part of the curve, if one assumes a fixed money wage

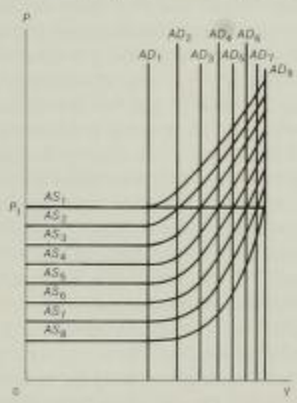
and a diminishing marginal product of labor. Firms will then offer a larger amount of output only at a price level higher by the amount needed to offset the diminishing marginal product of the additional labor that must be employed to produce that larger amount of output. As there is a range over which the marginal product of labor becomes smaller and smaller before it reaches zero, the assumption of a fixed money wage rate means that there is a range of the *AS* curve over which it becomes steeper and steeper before it reaches the vertical.[1]

On the assumption of a fixed money wage rate, the *AS* curve with an upward-sloping range is a closer approximation to reality than the one without such a range, and it is this form that will be employed hereafter. On the opposite assumption of a perfectly flexible money wage rate—and we will be returning to the classical theory in which this is a basic assumption—the *AS* curve is perfectly inelastic at the full-employment level of output.

On the other side of the analysis, the aggregate demand curve may take two forms: perfectly inelastic or less than perfectly inelastic, i.e., downward sloping to the right. In all of our work so far with the Keynesian model, we have assumed that the *AD* curve is perfectly inelastic. Like the perfectly elastic *AS* curve, this too is a simplification. If total nominal spending increases or decreases proportionally with any rise or fall in the price level, total real spending is unchanged. The same amount of goods is demanded whether the price level is higher or lower. There is some reason to expect this to occur if the real income received by resource suppliers as a group does not change with a change in the price level. For example, a doubling of prices makes goods twice as costly in nominal terms, but if income recipients as a group receive twice as much income in nominal terms as before, they may be expected to maintain real spending at the same level as before. This kind of conclusion is reasonable as far as it goes, but it omits various other ways in which a change in the price level may influence total real spending. These other influences work in a way to give the result that the aggregate quantity of goods demanded will be greater at a lower price level than at a higher price level or that the *AD* curve will slope downward to the right.

While it was a simplification of the Keynesian model to assume that the *AD* curve was perfectly inelastic, another simplification, this time in the classical model presented in Chapter 14, gave us an *AD* curve that did slope downward to the right without limit, specifically in the form of a rectangular hyperbola or specifically with unitary elasticity throughout. That *AD* curve was

[1] In terms of the present construction, *P* could remain stable all the way up to the full-employment level of *Y*, despite the diminishing marginal product of labor, if the money wage rate were to decrease by just the amount needed to offset the decrease in the marginal product of labor. Any one *AS* curve is drawn on the assumption of a given money wage rate. As shown in the figure, successive reductions in the money wage rate could shift the *AS* curve downward from AS_1 to AS_2 and so forth as needed to permit expansion of *Y* along a path with a stable price level of P_1.



Of course, there is no tendency for money wage rates to behave in any such fashion. On the contrary, the closer the economy gets to the full-employment level of output, the tighter labor markets become, and the greater becomes the upward pressure on the money wage rate. Wage rates tend to increase and thereby shift the *AS* curve upward and make the price level at which a given level of output is supplied higher than it otherwise would be.

based on nothing more than the simple quantity theory of money in which nominal spending is determined by the stock of money, given the theory's assumption of a stable velocity of money. As long as the stock of money remains unchanged, nominal spending remains unchanged. If M were \$100 and V a stable 5, nominal spending per time period would be established at \$500. A few points along the resulting unitary elasticity AD curve would be Y of 200 and P of 2.5, Y of 250 and P of 2, and Y of 300 and P of 1.67. In the simple quantity theory, this downward-sloping AD curve is fully specified by nothing more than the supply of money, given the assumption of a stable V. In turning now to the way in which a downward-sloping AD curve is derived from the IS and LM curves, we will find that, with the LM curve based on the demand for money that underlies the quantity theory of money, the AD curve will be a rectangular hyperbola or show unitary elasticity only over a limited range of output. With the LM curve based on the demand for money of the Keynesian theory, we will find a downward-sloping AD curve whose elasticity will ordinarily vary at each different price level.

DERIVATION OF THE AGGREGATE DEMAND CURVE AND DETERMINATION OF THE EQUILIBRIUM PRICE LEVEL AND OUTPUT LEVEL

In the IS-LM framework, the LM curve shows the various combinations of Y and r at which the real supply of money, m_s, and the real demand for money, m_d, are equal. As $m_s = M_s/P$, an increase in M_s is accompanied by a proportional increase in m_s as long as P is unchanged. Thus, on the assumption of an unchanged P, changes in M_s bring about shifts to the right in the LM curve in the way explained on pp. 297–99 and illustrated in Figure 16-8.

Unlike the case in those pages, P is here a variable. Now m_s can change either because of a change in M_s or in P or in both. For a numerical illustration, suppose that initially $M_s = 100$ and $P = 2$ so that $m_s = 50$. There will be a 20-percent increase in m_s, if there is a 20-percent increase in M_s with no change in P. We will have $m_s = 120/2 = 60$, a 20-percent increase over the initial m_s of 50. The same increase in m_s results with M_s unchanged at 100 but P reduced to 1.67 (or by 16.7 percent). That is, $m_s = 100/1.67 = 60$, a 20-percent increase over the initial m_s of 50. In terms of the LM curve, the same shift to the right will occur in both of these cases because in both cases there is the same increase in the real money supply.

In Part A of Figure 17-1 are IS and LM curves like those derived in the preceding chapter. The nominal money supply, M_s, remains unchanged in what follows and the initial price level of P_5 combined with this nominal money supply establishes the LM curve at LM_5. Successively lower price levels from P_4 through P_1 establish LM_4 through LM_1, respectively.[2] It is here assumed that the IS curve is unaffected by the price level, or its position is the same for all values of P shown. Suppose that the IS curve is that labeled IS_1 and that P is P_5. The equilibrium level of Y in Part A is at Y_1 given by the intersection of IS_1 and LM_5. Output of Y_1 will be demanded at the price level of P_5 as that is the price level underlying the LM curve which

[2]It is assumed that $P_5 - P_4 = P_4 - P_3 = P_3 - P_2 = P_2 - P_1$. This will be seen on the P axis in Part B of Figure 17-1. On this assumption, measuring on an arithmetic scale, the horizontal distance between LM_4 and LM_3 must exceed that between LM_5 and LM_4, that between LM_3 and LM_2 must exceed that between LM_4 and LM_3, and so forth. For example, suppose that M_s is given as 100. With $P = P_5$, $m_s = 100/5 = 20$. LM_5 indicates all combinations of r and y at which $m_s = 20 = m_d$. A decline in P to P_4 increases m_s to 25 and LM_4 indicates all combinations of r and y at which $m_s = 25 = m_d$. Apart from the liquidity-trap range, at each rate of interest the Y value indicated by LM_4 is 25 percent greater than that indicated by LM_5. With a further decline in P to P_3, $m_s = 33.33$ and LM_3 indicates all combinations of Y and r at which $m_s = 33.33 = m_d$. In this case, the Y value indicated by LM_3 at each interest rate above the liquidity trap is 33.33 percent greater

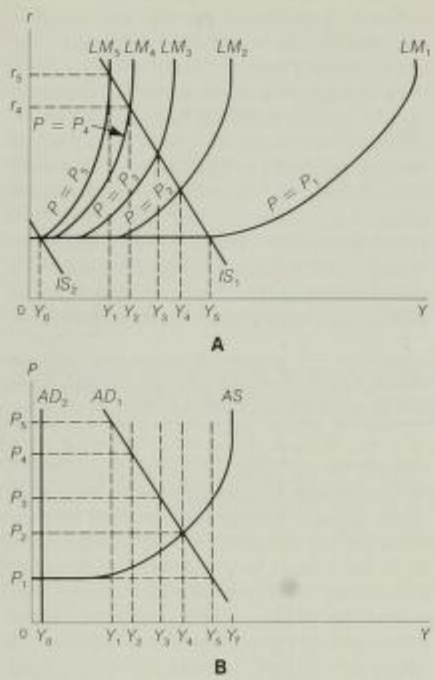


FIGURE 17-1
Equilibrium in the Keynesian Model

in combination with the *IS* curve identifies output of Y_1. Therefore, one point on the aggregate demand curve that is being derived in Part B must be at P_5 and Y_1. A decline in P to P_4 shifts the *LM* curve from LM_5 to LM_4 and establishes a new equilibrium level of Y in Part A at Y_2. As an output of Y_2 will be demanded at a P of P_4, a second point on the aggregate demand curve that is being derived must be at P_4 and Y_2. Further reductions in P to P_3, P_2, and P_1 shift the *LM* curve to LM_3, LM_2, and LM_1, respectively, and as before identify equilibrium levels of Y at Y_3, Y_4, and Y_5. These three levels of output will be demanded at P_3, P_2, and P_1, respectively, and thus establish three more points on the aggregate demand curve being derived. Connecting all the points now established, yields the downward-sloping aggregate demand curve labeled AD_1 in Part B.

What is the process by which a decrease in the price level results in an increase in the aggregate quantity of goods demanded? Consider an initial equilibrium in Part A at Y_1, r_5 given by LM_5 and IS_1. Assume then the decrease in P from P_5 to P_4 that shifts the *LM* curve from LM_5 to LM_4. The adjustment of Y and r in Part A to this change is not instantaneous so that, at the existing combination of Y_1, r_5, the public finds that it is holding larger real money balances than it wishes to hold. (Equilibrium between m_s and m_d is now given by the combinations of Y and r along the LM_2 curve, but the existing combination of Y_1, r_5 is to the left of the LM_2 curve.) The excess real balances spill over into the purchase of bonds, bid up bond prices, and lower the interest rate. As the interest rate falls, investment increases, and, other things being equal, the total amount of goods demanded increases. Once the adjustment is completed, the increase in the real money supply that results from the decline in P ceases to be an excess supply. The demand for money is $m_d = k(Y) + h(r)$. A fall in r to r_4 and a rise in Y to Y_2 increase m_d by just the amount of the increase in m_s, and there is again an equilibrium in the money market. There are any number of other combinations of Y and r along LM_4 that would also provide equilibrium in the money market, but all of these would be off IS_1. At any point off the *IS* curve, the amount

---

$(33.33/25 = 1.33)$ than that indicated by LM_4. In the same way, with a P of P_2, the Y value indicated by LM_2 at each interest rate is 50 percent greater than that indicated by LM_3. Finally, with a P of P_1, the Y value indicated by LM_1 at each interest rate is 100 percent greater than that indicated by LM_2. It may be noted that if the *LM* curves in Part A were to be drawn equidistant horizontally, the values of P must be such that $P_5 - P_4 > P_4 - P_3 > P_3 - P_2 > P_2 - P_1$. We will later have occasion to work with *LM* curves that are equidistant horizontally, but for our present purposes we want to work with equal absolute changes in P, and these give us the unequal spreads between the *LM* curves in Part A.

of goods demanded at the level of Y identified by that point is different from the amount of goods corresponding to that level of Y. To have simultaneous equilibrium in the money and goods markets, the new combination of Y and r must be Y_2, r_4. Therefore, while the aggregate amount of goods demanded was Y_1 with a P of P_5, a decline in P to P_4 sets into motion a process that gives us the result that the aggregate amount of goods demanded increases from Y_1 to Y_2.

Specifically how much of an increase in the amount of goods demanded will occur with the rightward shift in the LM curve produced by any particular decrease in the price level? The answer is found in the elasticity of the IS curve. The less elastic the IS curve, the smaller the increase in the amount of goods demanded that would result from any given decrease in P. If the IS curve were to reach the extreme of perfect inelasticity, the AD curve would also be perfectly inelastic. In this case the decline in r that would result from the fall in P would call forth no increase in investment and therefore no increase in the total amount of goods demanded. There is a special case in which the AD curve is perfectly inelastic, no matter how elastic the IS curve may be. The IS curve labeled IS_2 intersects the LM curves in the liquidity trap or the Keynesian range of the LM curve. Although declines in P shift the LM curve downward or rightward in the intermediate and classical ranges, they cannot do this in the Keynesian range. Therefore, such declines in P are powerless to reduce the interest rate below the rate found in the liquidity trap. As this means in turn that there can be no increase in investment via a lower interest rate and thus no movement down an elastic IS curve, an aggregate demand curve in Part B that is based upon an IS curve that intersects the LM curves in the Keynesian range will be perfectly inelastic, regardless of the elasticity of that IS curve. Thus, the perfectly inelastic AD_2 curve in Part B of Figure 17-1 is derived from the IS_2 curve and the LM curves in Part A.

Once an aggregate demand curve has been derived in Part B, whatever its position and whatever its elasticity, the intersection of that curve with the aggregate supply curve, assuming there is an intersection, determines the equilibrium level of output and the equilibrium price level. In the present illustration, these are Y_4 and P_2 with an IS of IS_1 and Y_0 and P_1 with an IS of IS_2. In all of our previous work with the Keynesian model, we employed an AS curve that was perfectly elastic up to the full-employment level of output, but as noted earlier, hereafter in working with the Keynesian model we will employ an AS curve with an upward-sloping range.[3] The particular AS curve shown in Part B of Figure 17-1 assumes that the fixed money wage rate is such that only at a price level of P_4 or higher will firms in the aggregate supply the amount of output, Y_f, that is produced with full employment of the labor force.

For the two illustrative AD curves derived in Part B of Figure 17-1, the equilibrium level of output is below the full-employment level. This follows our earlier work with the Keynesian model, which also showed that equilibrium could be below full employment. The difference now is that there is a range over which rightward shifts in the AD curve affect not only the output level as in the earlier Keynesian models but also the price level. A shift in the IS curve, for example, from IS_2 to IS_1 not only increases the equilibrium level of output but also raises the equilibrium price level.

While the Keynesian model shows that equilibrium below full employment is very much a possibility, it does not at all rule out that equilibrium may occur at full employment. In Part A of Figure 17-1, a sufficiently large rightward shift of the IS curve from the position at IS_1 will give rise to an AD curve in Part B that intersects the given AS curve at the full employment level of Y. However, while full employment may occur

[3]The derivation of this aggregate supply curve was described on pp. 219–23 of Chapter 13 and a review of these pages will be very helpful to an understanding of what follows.

in the Keynesian model, it need not necessarily occur in that model. The classical economists granted that it also need not occur in their model if the money wage rate was inflexible downward, but they insisted that it must occur if the money wage rate was perfectly flexible downward. As long as the money wage rate falls freely in the face of unemployment, it was their view that unemployment would be eliminated. While Keynes assumed a downwardly inflexible money wage rate in his model, an assumption that all agree is consistent with unemployment, he also asserted that the classical conclusion of automatic full employment does not necessarily follow even in a system in which the money wage rate is perfectly flexible downward. In the following section we will examine the way in which Keynes and the classical economists arrived at these different answers to a central question in macroeconomic theory: whether an economy characterized by downwardly flexible money wage rates and prices tends to automatically achieve equilibrium at the full-employment level of output.

WAGE–PRICE FLEXIBILITY AND FULL-EMPLOYMENT EQUILIBRIUM

The basis on which classical economics reached the conclusion that the equilibrium level of output would be only at the full-employment level was examined in Chapter 14 in some detail but in terms of a simple classical model. In that model, the aggregate demand curve is of unitary elasticity throughout as shown in Figure 17-2. This follows from the quantity theory of money with its assumption that the velocity of money is stable. Given V, MV or total spending is known as soon as M $(= M_s)$ is known. As $MV = PY$ and $AD = MV$, any indicated MV or AD will buy various quantities of Y depending on P. The various possible combinations of Y and P consistent with a given AD are, as noted, identified in the classical model by a curve with unitary elasticity that graphically appears as a rectangular hyperbola.

The classical theory, of course, assumes a flexible money wage rate, but as a first step let us here note briefly the result that follows from combining the aggregate demand curve of the basic classical model with the Keynesian assumption of a downwardly inflexible money wage rate.[4] In Figure 17-2, suppose that the rigid W is such that a P of P_4 gives the W/P at which S_n, the supply of labor, equals D_n, the demand for labor, or there is full employment. Therefore, at a P of P_4 or above, the AS curve is perfectly inelastic. However, with the downwardly inflexible W, at any P below P_4, W/P exceeds the W/P consistent with full employment, and therefore a lower level of employment is provided and a smaller aggregate quantity of goods is supplied by firms. In brief, the AS curve slopes downward to the left below P_4 as shown by the broken-line AS curve in Figure 17-2. With the nominal money supply being such as to yield the AD curve shown in this figure, the result is an intersection between the AD curve and the broken-line AS curve at a Y below Y_f. In the illustration, full employment requires a Y of 200; actual Y is 160. The conclusion: With a rigid money wage rate, equilibrium can occur in the classical model at an output level below the full-employment level.

However, once we return to the classical assumption that the money wage rate falls freely any time there is unemployment, the AS curve that emerges is a perfectly inelastic curve at the full-employment level of output as shown by the solid-line AS curve in Figure 17-2.[5] With

[4] For a fuller development of the brief statement provided here, see Chapter 14, pp. 241–43.

[5] The derivation of this aggregate supply curve was described on pp. 223–25 of Chapter 13, and, as with the derivation of the upward-sloping supply curve in Figure

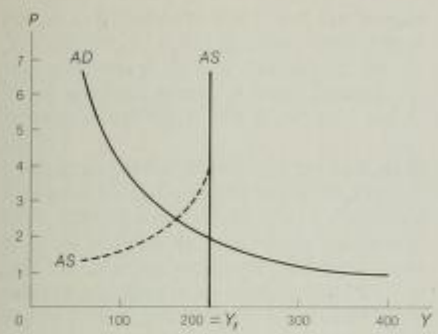


FIGURE 17-2
Full-employment equilibrium in the basic classical model

an *AD* curve of the kind given by the simple quantity theory of money, there must necessarily be an intersection between that curve and the perfectly inelastic *AS* curve, and the level of output at which this intersection occurs is necessarily the full-employment level. In the illustration, Y_f is 200 and P is 2. Of course, if the price level were above 2, say at 4, and if for some reason the price level were inflexible downward, the existing total spending would be unable to purchase the full-employment output of 200. The amount purchased would be 100; reduced production and unemployment would follow. However, in a system of competitive markets for labor and goods, the money wage rate will fall freely in the face of unemployment and the pursuit of profits by competitive sellers can be depended on to lead to a falling price level also. Given the assumptions of the classical model, the system will move to the full-employment equilibrium indicated by the intersection between the *AD* curve and solid-line *AS* curve.

17-1, a review of the pages on the derivation of the perfectly inelastic supply curve will be very helpful to an understanding of what follows.

To combine the simple quantity theory of money with a downwardly flexible money wage rate and price level is to make the full-employment equilibrium shown in Figure 17-2 inevitable. Thus, Keynes could not accept a downwardly flexible money wage rate and still reject the classical conclusion unless he rejected the simple quantity theory of money in which total real spending is directly and proportionally related to the real money supply. Reject this he did. But how does the rejection of the quantity theory of money get one to Keynes's conclusion that output may be below the full-employment level even with a downwardly flexible money wage rate? A necessary first step to an answer is to trace the way the *AD* curve is derived from the *IS-LM* curves for the case in which the *LM* curves are the kind that result with the quantity theory of money as part of the model.

Keynes made the demand for money a function of both the income level and the interest rate or $M_d/P = m_d = k(Y) + h(r)$. The derivation of the *LM* curve for this money-demand function was shown in Figure 16-4 on p. 290. If the demand for money is interest-elastic as in the Keynesian function, there is a range of the *LM* curve over which it slopes upward to the right. In the simple quantity theory, the demand for money is not interest-elastic. The quantity theory suppresses the speculative demand function for money, $M_{sp}/P = h(r)$, or makes the demand equation $M_d/P = m_d = k(Y)$. For a given price level, the result is that the *LM* curve based on $m_d = k(Y)$ is a vertical or perfectly inelastic line with respect to the interest rate as shown in Part D of Figure 17-3. Unlike Figure 16-4, m_{sp} in Part A of Figure 17-3 is zero at all interest rates and the m_{sp} curve disappears. In Part B, whatever m_s happens to be, it is all necessarily held in transactions balances, m_t. Running a horizontal line from the point on the m_t axis at which $m_t = m_s$ across to the $m_t = k(Y)$ function in Part C identifies the level of Y at which $m_s = k(Y)$. It is the level of Y so identified at which $m_s = m_d$, whatever the interest rate, as $m_d = m_t$ and $m_t = m_s$. The *LM* curve be-

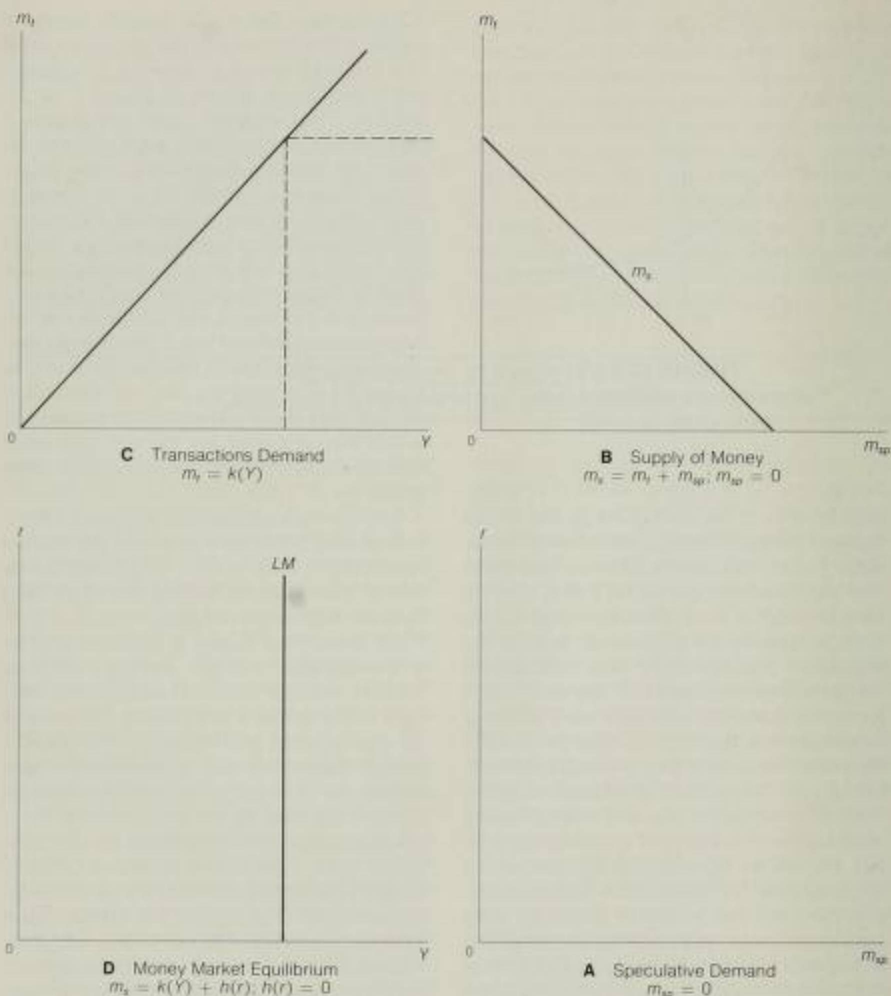


FIGURE 17-3
Money market equilibrium with speculative demand equal to zero

comes a perfectly inelastic line at that level of Y at which $m_s = k(Y)$.[6] Its position will change with a change in m_s or a change in k, but the curve itself will remain perfectly inelastic. Its position along the Y axis is simply equal to m_s/k.

Assume now that M_s/P and k are such as to locate the LM curve at LM_5 in Part A of Figure 17-4. With no change in M_s or in k, decreases in P will increase the real money supply and shift the LM curve to the right. This is the same relationship noted in connection with Part A of Figure 17-1. As in that figure, the present figure shows five LM curves corresponding to five levels of P. Also as before, the intersections with an IS curve identify five points in Part B which are connected to form the AD curve. Two IS curves, IS_1 and IS_2, are shown in Part A, but both yield the same AD curve in Part B as both intersect each of the series of LM curves at the same level of Y. The AD curve derived in Part B is of unitary elasticity *over the range shown* in the figure. As with the AD curve in Figure 17-2 derived directly from the simple quantity theory of money, the AD curve in Figure 17-4 shows that output may be raised from a less than full-employment level such as Y_3 to the full-employment level of Y_4 by a decline in the price level from P_3 to P_2. With the AD and AS curves as shown in Part B, unemployment will necessarily be eliminated by an appropriate decline in the price level.

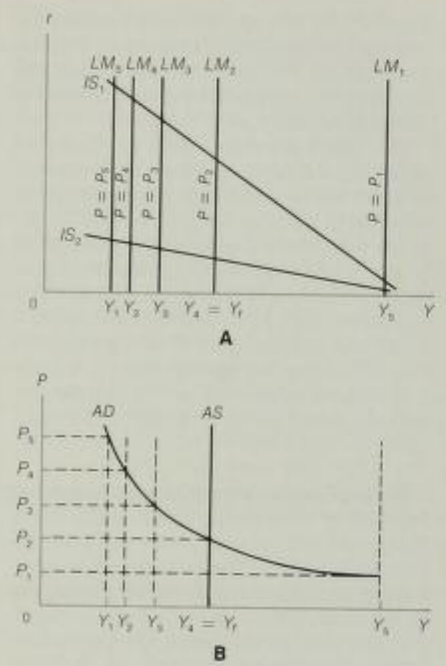


FIGURE 17-4
Full-employment equilibrium in the classical model within the IS-LM framework

Wage–Price Flexibility and the Interest-Rate Effect

In the simple quantity theory on which Figure 17-2 is based, an increase in the real money supply that results from a reduction in the price level leads *directly* to an increase in the quantity of goods demanded and successive increases in the real money supply lead directly to successive increases in the quantity of goods demanded. The relationship is different in Figure 17-4. There the increase in the real money supply does not lead directly to an increase in the quantity of goods demanded but achieves that result indirectly, specifically by first effecting a decline in the interest rate. In Part B of Figure 17-4, the increase in the quantity of goods demanded that occurs with a decline in P from P_3 to P_2 is explained in Part A not merely through the shift in the LM curve from LM_3 to LM_2, that being the result of the simple quantity theory, but also through the fact that this re-

[6] The real money supply, M_s/P, is also assumed to be interest-inelastic in this model or it would provide a slope to the LM curve despite the interest inelasticity on the demand side.

duces the interest rate and thereby raises the quantity of capital goods demanded. This rise in the quantity of capital goods demanded is the source of the increase in the total quantity of goods demanded from Y_3 to Y_4.

The simple quantity theory of money on which the *AD* curve of Figure 17-2 is based does not require a reduction in the interest rate as a part of the process by which an increase in the real money supply brings about an increase in the quantity of goods demanded. However, such a reduction is required in the Keynesian theory on which Figure 17-4 is based and this difference in turn brings us to one of the bases for the Keynesian conclusion that full employment may be unattainable even with wages and prices perfectly flexible downward.[7]

Inconsistency between Saving and Investment In Figure 17-4, assuming initially that output is below the full-employment level, that level may be attained via a decline in the price level to P_2, but that result requires that the *IS* curve be positioned far enough to the right, e.g., like the IS_1 curve, or alternatively that the *IS* curve be sufficiently interest-elastic, e.g., like the IS_2 curve. Instead of either of the *IS* curves in Part A of Figure 17-4, suppose that the *IS* curve is that shown in Figure 17-5. Referring back to Figure 16-7 on p. 296, it will be seen that the position of the *IS* curve depends on the position of the investment curve in Part A and the position of the saving curve in Part C. A leftward shift of the investment curve or the saving curve will cause a leftward shift of the *IS* curve. If the position of the investment and saving curves are such as to yield an *IS* curve in the relatively low position of that in Part A of Figure 17-5, the result is that declines in the price level are unable to restore to full employment a system suffering from unemployment.

[7]These different views of the process by which a change in the real money supply is transmitted through the system to affect the real income level are considered further in Chapter 25, pp. 506–535.

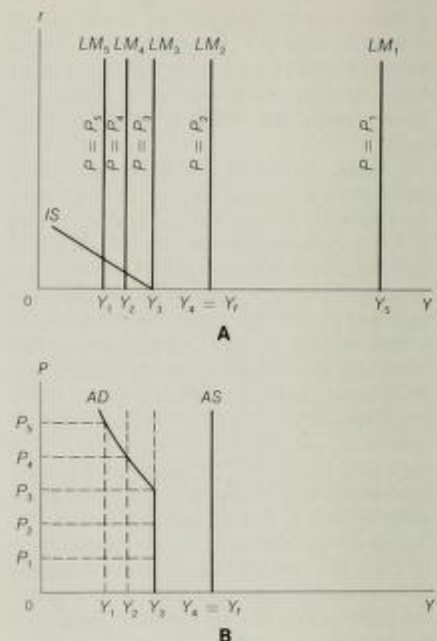


FIGURE 17-5
Unemployment due to inconsistency between saving and investment

To see this, start off in Part A with *P* of P_5 which, given the nominal money supply and the value of *k*, establishes the *LM* curve at LM_5. As in Figure 17-4, the quantity of goods demanded in Part B with a *P* of P_5 is found to be Y_1, a quantity well below Y_4, the full-employment level of output in this illustration. There is unemployment. On the assumption of flexible wages and prices, the money wage rate falls and the price level falls in turn. As *P* falls step by step over the range from P_5 to P_3, the quantity of goods demanded increases as shown by the portion of the *AD* curve in Part B derived from that range of price levels in Part A. Over this particular range, the *AD* curve is of unitary elas-

ticity as in the preceding figure. However, once P falls below P_3, the AD curve in the present figure becomes perfectly inelastic. Such decreases in P bring about no further increases in the quantity of goods demanded. The quantity demanded reaches a maximum of Y_3, short of the quantity, Y_4, produced at full employment. Thus, the conclusion: Full employment is not attainable through a falling price level.

If the increase in the real money supply effected by a decrease in the price level always led to an increase in real demand proportional to the increase in the real money supply, a result like this could not occur. However, in Part A of Figure 17-5 an increase in the real money supply affects real demand only by first reducing the interest rate and boosting investment. For the simple case of a two-sector economy, even if the interest rate could be reduced to zero, the amount of investment called forth at a zero interest rate would be equal to the amount of saving at Y of Y_3, as this is the level of Y at which the IS curve in Part B hits the Y axis. To achieve a Y of Y_4 would require that an interest rate of zero call forth an amount of investment equal to what saving would be at Y_4. Then the IS curve would hit the Y axis at Y_4. However, the underlying saving and investment curves give us the IS curve as shown and the maximum attainable output level is the less than full-employment level of Y_3.[8]

It is an open question whether the relationship between the investment schedule and the saving schedule could be such that the amount of saving generated at full-employment output would exceed the amount of investment generated by a zero interest rate. But it must be recognized as a possibility whose occurrence precludes the attainment of the full-employment output through a deflation of wages and prices.

[8] It appears from the diagram that a negative interest rate could get the output level to Y_4, but a negative rate cannot be secured. While the interest rate might approach zero, it surely could not become negative (in which event lenders would be required to pay interest to borrowers to use the lenders' funds).

Liquidity Trap There is a second basis for the Keynesian conclusion that downward flexibility of wages and prices may fail to correct unemployment. On this basis, one need not deny that a sufficiently low positive interest rate is able to call forth investment equal to what saving will be at full employment. What one does deny on this other basis is that the interest rate can fall to the level at which this occurs. What is now brought into the picture is the concept of the liquidity trap, which sets a lower limit to the decline in the interest rate. The perfectly inelastic LM curves of Part A of Figure 17-5 follow from the demand for money of the simple quantity theory in which $M_d/P = k(Y)$. What we now have in their place are the LM curves of Part A of Figure 17-6 which follow from the demand for money of Keynes's theory in which $M_d/P = k(Y) + h(r)$ and in which at a sufficiently low rate of r the speculative component of the demand for money becomes perfectly elastic.

As before, assume that P is initially P_5, which places the LM curve at LM_5. The LM_5 curve intersects the IS curve at Y_1 and gives us a point on the AD curve in Part B at P_5 and Y_1. For P_4, P_3, P_2, and P_1, other points on the AD curve are identified as before. With each decline in P from P_5 to P_1, there is an increase in the quantity of goods demanded. However, with P at P_1, the quantity demanded, Y_5, is still below Y_f. What apparently is needed is still more deflation to increase the quantity of goods demanded. However, the deflation-generated increase in the real money stock has already pushed the interest rate down to r_1, which in the present illustration is the minimum rate set by the liquidity trap. To get an increase in the quantity of goods demanded beyond Y_5 requires an increase in investment, which in turn requires a decline in the interest rate. However, further declines in P below P_1 and the resulting further rightward shifts in the LM curve beyond LM_1 are unable to reduce the interest rate below r_1. As they call forth no increase in investment, declines in P below P_1 leave the total quantity

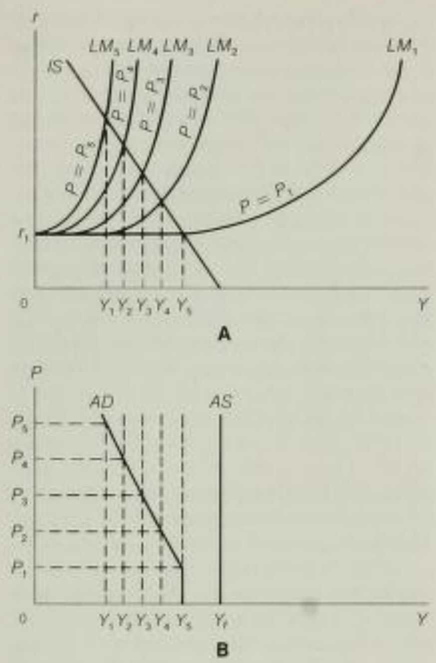


FIGURE 17-6
Unemployment due to a liquidity trap

of goods demanded at Y_5 or the *AD* curve becomes perfectly inelastic below P_1. In this illustration, the quantity of goods demanded reaches a maximum of Y_5, short of the quantity produced at full employment, Y_f. Thus, the conclusion: Full employment is not attainable through a falling price level.

Keynes's conclusion that a deflation of wages and prices could fail to produce the result asserted by the classical economists rests upon his view of the way that deflation affects the quantity of goods demanded. As he saw it, whatever increase there was in the quantity of goods demanded at a lower price level was essentially the result of the lower interest rate that results from a lower price level. There are other ways in which a decline in the price level may affect the quantity of goods demanded but this interest rate effect so emphasized by Keynes has come to be associated with his name and is sometimes referred to as the Keynes effect.

A Policy Question: Price Deflation versus Nominal Money Supply Expansion We have now reviewed the bases on which Keynes argued that deflation may be unable to get the system to full employment: The *IS* curve may lie so far to the right that not even a zero interest rate will be sufficient, or the liquidity trap may prevent the needed decline in the interest rate which otherwise would be sufficient. However, apart from the possibility that wage and price deflation may be unable to get output to the full-employment level no matter how great the fall in prices, Keynes and his followers argued that, as a policy matter, there is in any event little to be said for trying to get to full employment in this way. If one accepts Keynes's position that whatever favorable effect on output follows from wage and price cuts comes through their effect on the interest rate, then to achieve a reduction in the interest rate through the painful process of deflation is "patently absurd," as the late Professor Hansen, Keynes's foremost American disciple, once put it. As we have seen, the shift in the *LM* function necessary to reduce the interest rate may be brought about by an increase in the money stock or by a decrease in the price level. The objective of reducing the interest rate can be achieved much more simply through an increase in the money supply. This can be done by the central bank without any direct interference with the wage and price structure. Working toward the same objective through wage and price cuts is to do it the "hard way." It was for this reason that Keynes identified wage cuts as "monetary management by the Trade Unions, aimed at full employment."[9]

[9]John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, Harcourt Brace Jovanovich, 1936, p. 267.

Leaving monetary management to the banking system also avoids many difficult problems that arise in achieving the goal of a lower interest rate through wage and price reductions. In the first place, one must consider all the institutional barriers that stand in the way of a fall in wages and prices. One must also consider all the economic inequities and distress that inevitably result from wage and price cuts. For all wages and prices do not fall at the same rate; some groups benefit and others suffer as income and wealth are redistributed in the course of a deflation. Businesses that are saddled with heavy fixed debt may be unable to weather the storm; bankruptcies are bound to result. Most important of all, because in practice a decline in wages and prices does not occur in one quick step, one wage cut and one price cut often give rise to expectations of further cuts, and such "bearish" expectations will lead to a postponement of some investment spending (and probably also some consumption spending on durable goods). In this fashion, the favorable effect on the quantity of goods demanded, which otherwise occurs in response to the lower interest rate that results from the reduction in the price level, could easily be swamped by the unfavorable effect on the quantity of goods demanded of the decline in spending that results from expectations of further declines in wages and prices. For these and other reasons, most economists agree that if a falling price level affects the quantity of goods demanded through its effect on the interest rate, an expansionary monetary policy is assuredly to be preferred to a falling price level as the means of achieving the goal of a lower interest rate and higher output and employment levels.

Wage–Price Flexibility and the Pigou Effect

As we have seen, if the limit to the reduction in the interest rate set by the liquidity trap is reached before the quantity of goods demanded can be raised to the full-employment level or if even a reduction in the interest rate to zero is unable to raise the quantity of goods demanded to the full-employment level, wage and price flexibility working through the Keynes effect is repudiated as a means of achieving the full-employment level of output. However, this conclusion is based on the argument that a decline in wages and prices exerts its influence only through the interest rate. An attempt to counter the Keynesian argument and rehabilitate the classical theory's conclusion of automatic full employment through wage–price flexibility is found in the Pigou effect or the real balance effect, first advanced by Professor A. C. Pigou in the early forties and since then the subject of a sizable literature.[10]

How does the Pigou effect work? Suppose that investment falls off so that output and employment drop as in the ordinary Keynesian model. Unemployment causes money wages to fall, which means lower costs and lower prices. The prices of assets such as goods, buildings, land, and common-stock shares may be expected to fall with other prices so that there will be no change in their real value. However, the fall in the price level means a rise in the real value of assets that are fixed in dollar terms, such as money, savings deposits, and bonds. This increase in the real value of fixed-dollar assets makes the holders less anxious to build up further their asset holdings. They devote a smaller fraction of their current income to saving and a larger fraction to consumption, which amounts to a downward shift in the saving func-

[10] A. C. Pigou, "The Classical Stationary State," in *Economic Journal*, Dec. 1943, pp. 345–51, and "Economic Progress in a Stable Environment," in *Economica*, Aug. 1947, pp. 180–88, reprinted in *Readings in Monetary Theory*, pp. 241–51; and D. Patinkin, "Price Flexibility and Full Employment," in *American Economic Review*, Sept. 1948, pp. 543–64, also reprinted in *Readings in Monetary Theory*, pp. 252–83. See also D. Patinkin, *Money, Interest, and Prices*, 2nd ed., Harper & Row, 1965, and B. P. Pesek and T. R. Saving, *Money, Wealth, and Economic Theory*, Macmillan, 1967. As is commonly done, the terms, Pigou effect and real-balance effect, are here used as synonyms, although this is not altogether proper. For an explanation of the difference between them, see G. E. Makinen, *Money, The Price Level, and Interest Rates*, Prentice-Hall, 1977, pp. 160–62.

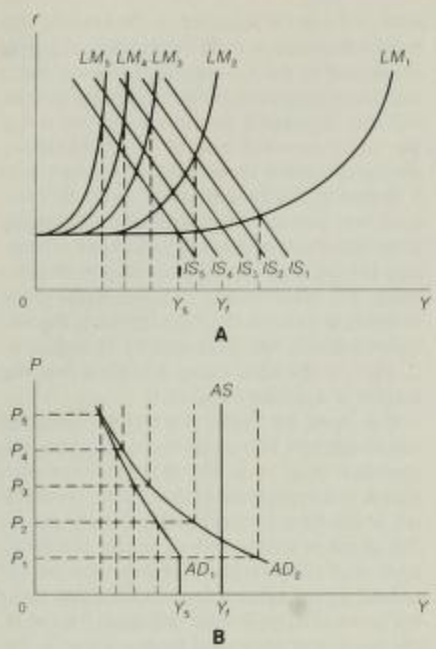


FIGURE 17-7
Full employment via the Pigou effect

tion or an upward shift in the consumption function. In terms of Figure 17-7, this appears as a rightward shift in the *IS* function. With each successive drop in *P* from P_5 to P_4 and so forth, the *IS* curve shifts from IS_5 to IS_4 and so forth. The *LM* curves and the IS_5 curve in Figure 17-7 are the same as the *LM* curves and the *IS* curve in Figure 17-6, and the AD_1 curve derived from them shows that the liquidity trap sets an upper limit of Y_5 to the quantity of goods demanded. However, with the Pigou effect allowed for, the *IS* curve as well as the *LM* curve now shifts to the right with each decline in *P*, and the *AD* curve in Figure 17-7 shifts from AD_1 to AD_2. The AD_2 curve is derived from the intersections of IS_5 and LM_5, IS_4 and LM_4, and so forth. Because of the Pigou effect, the quantity of goods demanded is now higher than before at each price level below P_5. The AD_2 curve intersects the *AS* curve and therefore yields equilibrium at the full-employment level of output. Thus, according to this argument, there is some decline in the price level which will raise the real value of a given stock of fixed dollar assets sufficiently to shift the saving function downward by the amount necessary to shift the *IS* curve so as to produce an *AD* curve consistent with the full-employment level of output.

We mentioned the Pigou effect in Chapter 9 in our discussion of factors influencing consumption spending. In terms of the consumption function, the Pigou effect makes consumption depend not only on the level of real income (and other factors), but also on the real value of consumers' holdings of fixed-dollar assets. If wages and prices fall proportionally, as Keynes argued they do, real income remains unchanged and so, to the extent that real consumption depends on real income, consumption remains unchanged. But if consumption also depends on the real value of consumer holdings of fixed-dollar assets, we may expect a higher average propensity to consume at each level of real income as a result of a lower price level. This means an upward shift in the consumption function, and this in turn means a rightward shift of the *IS* function and from this a rightward shift in the *AD* function. With the Pigou effect included in the model, there is presumably a decrease in wages and prices sufficient to shift the *AD* function rightward to the position at which the quantity of goods demanded at the new lower price level will be equal to the full-employment output of goods.

To the extent that the Pigou effect can shift the *AD* curve in this way, it may be said that Pigou met Keynes on the latter's own ground and there, at least in terms of pure theory, won a "triumph" for classical theory by showing that, if flexible wages and prices are incorporated, even the Keynesian model gives us the automatic full-employment conclusion of classical theory. This seems to invalidate the Keynesian position that such flexibility can not yield the classical conclusion. But economists

in general concede no such victory to Pigou and the classical theory. They raise a number of criticisms of the Pigou effect and of the attempt to rehabilitate classical theory through its use.

A first criticism is that the Pigovian argument cannot apply, in the simple form given above, to all fixed-dollar assets. Although the real value of all such assets held by a creditor increases as the price level falls, the real value of the corresponding debtor's obligations also increases. Conceivably the rise in the creditor's average propensity to consume could be just offset by the fall in the debtor's, leaving the *APC* of the system as a whole unchanged. The Pigovian reply to this attack is to limit the argument to fixed-dollar obligations issued by the government, on the reasonable assumption that this debtor's expenditures will not be affected one way or the other by changes in the real burden of its debt resulting from changes in the price level. In other words, deflation increases the real value of government debt holdings, the sum of currency and interest-bearing government securities, and thus stimulates consumption spending by those who hold these assets, without at the same time being offset by any decrease in government spending. The net effect, as argued by Pigou, is an increase in aggregate spending.

However, even if it is granted that a deflation-created rise in the real value of consumer holdings of currency and government securities will increase consumption spending at any given level of real income, the crucial question then becomes how great the rise in consumer spending will be for any given decrease in the price level. In the face of a certain amount of unemployment, it is one thing if a 10-percent decline in the price level and the accompanying increase in the real value of the stock of currency and government securities is sufficient to raise consumption spending by the amount needed to restore full employment; it is quite another if the Pigou effect is so weak that the same result can be achieved only with an 80-percent decline in the price level. If a major deflation is required, this in itself effectively rules out reliance on the Pigou effect as a practical means of restoring an economy to full employment.[11] A hyperdeflation may satisfy the purely theoretical requirements of the Pigou effect, but in practice it might also produce economic distress leading to riots and even revolution.

If anything more than a minor deflation is required for the Pigou effect to do its work—and this seems to be the case—we have the final and decisive objection to the Pigovian theory: it neglects the role played by expectations. This is the same objection we noted earlier in discussing deflation as a means of reducing the interest rate. An instantaneous once-and-for-all deflation needed to give credibility to the Pigou effect simply cannot be obtained in practice. Wages and prices cannot be substantially deflated without creating expectations of further deflation. An initial fall in wages and prices will often generate pessimistic forecasts and cause business persons and consumers to reduce their expenditures.

It may even be argued that a once-and-for-all deflation would not work, assuming that such a thing could be engineered. The public might well view the lower price level as a temporary situation to be reversed before long. A return to the earlier higher price level may be expected as unemployment declines and the economy moves into a recovery. This means that the public will also view the increase in the real value of their fixed-dollar assets as a temporary increase. While a permanent increase in the value of these assets is another matter, it is doubtful that such a temporary increase will raise expenditures above the level they would otherwise attain.

[11]Empirical testing has given mixed results. That the Pigou effect is too weak to have any practical significance is suggested by T. Mayer's study, "The Empirical Significance of the Real Balance Effect," in *Quarterly Journal of Economics*, May 1959, pp. 275–91. A study that indicates a Pigou effect of considerable strength is Ta-Chung Liu, "An Exploratory Quarterly Econometric Model of Effective Demand in the Postwar U.S. Economy," in *Econometrica*, July 1963, pp. 310–48, and especially pp. 331–32. See also D. Patinkin's discussion of some other investigations in *Money, Interest, and Prices*, pp. 651–64.

A final question is whether or not it is realistic to believe that the effects of a lower price level on private creditors and private debtors cancel out in the way suggested earlier. Is the increase in spending that we may expect from creditors in response to a rise in the real value of their fixed-dollar assets likely to be as large as the decrease in spending that we may expect from business and personal debtors in response to the rise in the real burden of their fixed-dollar debts? Short of resorting to bankruptcy, there is no escape from the increase in this real burden—other things being equal, the real income of the debtors remaining after paying interest is reduced and their real spending is fairly likely to change in line. It is doubtful that creditors whose real income has been increased, at least temporarily, will increase their real spending in an offsetting manner. There is thus real pressure on debtors to cut back but no corresponding pressure on creditors to do the opposite. The net effect may well be a reduction in private spending.

All things considered, the Pigovian argument appears impractical as an approach to the solution of the unemployment problem. In fact, Professor Pigou himself did not really see it being used for this purpose. He described certain aspects of his argument as "academic exercises, of some slight use for clarifying thought, but with very little chance of every being posed on the chequer board of actual life."[12]

Wage–Price Flexibility and Other Effects

The Keynes (interest-rate) effect and the Pigou (real-balance) effect have received the most attention in discussions of wage-price flexibility as a cure for unemployment, but there are several other effects that deserve mention.

Income-Redistribution Effects Wage and price deflation involves some redistribution of real income in favor of fixed income groups. For any level of real income, the share of the total that goes to wage and profit recipients will decrease, and the share going to recipients of interest, rents, and pensions will increase. Some fixed income flows such as interest payments may go predominantly to upper-income groups, and others such as pensions may go predominantly to lower-income groups, but knowledge of the income distribution of these total flows is limited. To the extent that redistribution is in favor of lower-income groups, some rise in the consumption function is to be expected, and therefore so is some shift to the right in the *IS* function and the *AD* function. But the extent of this effect depends in turn on whether there is a sizable difference between upper- and lower-income groups in the marginal propensity to consume. Recall from our discussion in Chapter 9 of income distribution as a factor influencing consumption spending that the difference between the MPCs at different levels of family income is not substantial. Since we cannot be sure whether redistribution is in favor of lower- or upper-income groups, and since the difference in the MPCs is not appreciable in the first place, we can do no more than note the existence of fixed incomes as a factor to be considered in appraising the effects of wage–price declines on consumption spending. However, this factor, in any event, is probably of no great importance except during a drastic deflation.

Tax and Transfer-Payment Effects Through their tax effects, wage–price declines may be expected to have a favorable effect on consumption. With the progressive income tax the mainstay of the federal revenue system, deflation automatically shifts taxpayers into lower brackets and reduces the fraction of their real income that is paid in income taxes. There is then an increase in real disposable income and an increase in consumption. To some extent this favorable effect is offset by the existence of specific taxes (such as 10¢ per pack of cigarettes) the burden of which increases with defla-

[12] F. A. Lutz and L. W. Mints, eds., *Readings in Monetary Theory*, p. 251.

tion. The net effect is favorable, however, in a system whose overall tax structure is progressive. In those government programs that call for purchase of a fixed quantity of goods and services, expenditures on goods and services will tend to decline in line with the fall in price level. However, since some transfer payments are at any time fixed in dollar terms, a fall in the price level means a rise in the real income represented by a given dollar flow of these transfer payments. Therefore, we may expect an increase in real consumption by the beneficiaries of such payments.

Notice that government could provide the same stimulus to consumption spending, with an unchanged price level, by reducing taxes and raising transfer payments. The situation is comparable to that discussed earlier in connection with the Keynes effect. Just as an expansionary monetary policy is an alternative to deflation as a means of reducing the interest rate, so an expansionary fiscal policy of cutting taxes and raising transfer payments is an alternative to deflation as a means of increasing after-tax incomes and thereby raising consumption expenditures. But, referring specifically to the wage–price flexibility argument, if there is deflation, tax and transfer-payment effects must be recognized as another stimulus to aggregate spending.

Foreign Trade Effects The final effect to consider is that on a nation's imports and exports. A decline in a nation's prices relative to the level of prices in other nations encourages exports and discourages imports, increasing the net export (or decreasing the net import) component of aggregate spending. In an open economy, aggregate spending is the sum of $C + I + G + (X - M)$, and a rise in $(X - M)$ will shift the *IS* function to the right, just as a rise in C, I, or G would. Because imports and exports play a much more important role in countries like England than in a country like the United States, the foreign trade effect is more important there than here, but in an open economy some stimulus to demand and a shift in the *IS* function may be expected from a decline in its price level relative to the price level in other countries.

In summary, classical theory argued that an economy with flexible wages and prices would be self-equilibrating at the full-employment income level. We have examined this argument in terms of the various effects of wage–price flexibility: the Keynes effect (or interest-rate effect), the Pigou effect (or real-balance effect), and the effects of income redistribution, tax and transfer payments, and foreign trade, each of which covers one of a number of ways in which wage–price flexibility can produce changes in the level of output. Keynes's position that deflation affects the output level through a shift in the *LM* function and a reduction in the interest rate amounts to a repudiation of deflation as a sure road to the full-employment level of output. All the other effects operate not through shifts in the *LM* function but through shifts in the *IS* function. Output-expanding shifts in the *IS* function as a result of deflation are limited by no barrier such as the liquidity trap. A sufficient decline of wages and prices might be capable of doing what classical theory said it would do—restore the economy to full employment—but not in the manner indicated by the crude quantity-theory reasoning examined in Chapter 14. Furthermore, the whole wage–price flexibility argument holds up only if the required deflation occurs without creating widespread expectations of further deflation during its course.

Although one can thus find some theoretical basis for the conclusion that an economy suffering general unemployment may find a cure through flexible wages and prices, it does not necessarily follow that this should be the medicine prescribed. While deflation may possibly bring the patient around, so too may other measures. Fiscal policy has been the most popular of these since the appearance of the *General Theory*. Beyond fiscal policy and, of course, monetary policy lie other measures, which reach an extreme with the outright socialization of investment. Whatever the appropriate policy for any set of circumstances, the clas-

sical policy of wage–price flexibility is unlikely ever again to receive serious practical consideration, certainly not after the revolution in economic theory of the past three decades. It is now almost inconceivable that government would stand by passively in the face of deflation

> to test the proposition that if a deflating economy is left alone long enough, it will eventually, via a deflationary wage and price spiral, dig itself deeply enough into the mire to extricate itself. It is a political axiom of the mid-twentieth century that a government that conscientiously pursued this policy would not be around to observe the outcome.[13]

Today it is virtually certain that there will be some, even massive, intervention to check and reverse a rapid deflation that might conceivably occur in a sharp and sudden business downturn. Although this would be the public policy pursued, it is important to note here that the *theoretical* support for reliance on the classical solution of flexible wages and prices was not so completely swept away as Keynes and his disciples believed it was in the early days following the appearance of the *General Theory*. This question is still disputed by Keynesian and anti-Keynesian theorists, but it has now become a question of pure theory removed from the province of applied economics. Few economists today, however anti-Keynesian, would prescribe the classical theory's medicine as a cure for general unemployment. However, the reason is not that it will necessarily fail to effect a cure, but that to most it is less palatable than the medicines of expansionary monetary and fiscal policies.

MONETARY–FISCAL POLICIES AND THE FULL-EMPLOYMENT EQUILIBRIUM

At the end of the preceding chapter we looked briefly at the way in which monetary and fiscal policies shift the *LM* and *IS* curves to affect the output level in a model with a stable price level. It was there assumed that the aggregate supply curve was perfectly elastic up to the full-employment level and that the economy was operating below the full-employment level so that shifts in the aggregate demand curve could affect the output level but not the price level. We here replace the perfectly elastic aggregate supply curve with one that slopes upward to the right. The result is that shifts in the aggregate demand curve produced by shifts in *LM* and *IS* curves affect not only the output level but also the price level. As the position of the *IS* and *LM* curves are in turn affected by changes in the price level, to trace the process by which monetary and fiscal policies raise the level of output requires that allowance be made for the changes in the price level that may occur as a result of those policies. This is something that can be done here but was ruled out in the preceding chapter by the assumption that the economy was operating along the perfectly elastic portion of the aggregate supply curve.

In Part A of Figure 17-8, the three solid-line *LM* curves are those for a given nominal money supply, M_s, and three different price levels, P_3, P_2, and P_1.[14] A decrease in P is an increase in the real money supply, and, other things being equal, an increase in the real money supply

[13] J. P. Lewis and R. C. Turner, *Business Conditions Analysis*, 2nd ed., McGraw-Hill, 1967, p. 288.

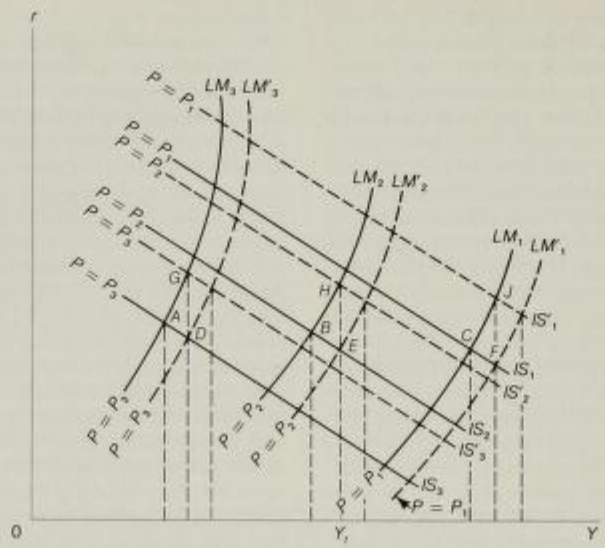
[14] Unlike the assumption made in drawing the figures in the preceding section, here it is *not* assumed that $P_3 - P_2 = P_2 - P_1$, a condition which, as we saw there, requires that the spread between the *LM* curves increase with each decrease in P. Here the spread between the LM curves is the same but $P_3 - P_2 > P_2 - P_1$. Also, here the *LM* curves are not drawn to show a Keynesian range and a classical range. The inclusion of these properties is not needed for present purposes and would needlessly complicate the diagram.

shifts the *LM* curve to the right. The *LM* curve will also shift to the right with an increase in M_s and an unchanged *P*. Thus, with *P* given at P_3, there is some increase in M_s, which will shift the *LM* curve from LM_3 to the broken-line LM'_3. Similarly, this same increase in M_s accompanied by a price level below P_3 will shift the *LM* curve rightward from its previous position. With a *P* of P_2, the position will be that of LM'_2 instead of LM_2; and with *P* of P_1, the position will be that of LM'_1 instead of LM_1. All this says is that for each M_s there is a different *LM* curve for each *P*. In Part A, there are two amounts of M_s and three levels of *P* or two sets of three curves each. To work, for example, with three different amounts of M_s and five levels of *P* would require three sets of five curves each (some of which might coincide).

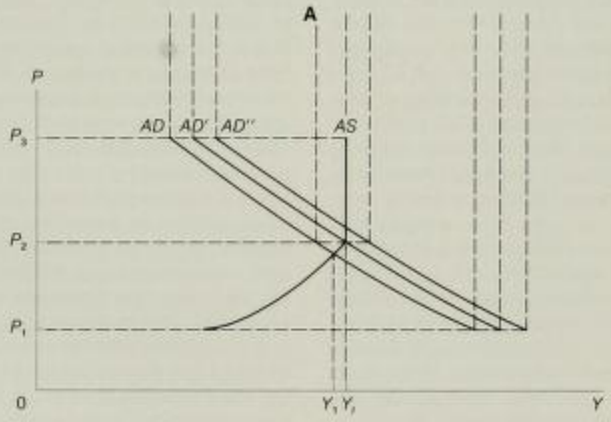
Part A of the figure also includes a set of *IS* curves similar to the set of *LM* curves. The three solid-line *IS* curves are the original curves corresponding to the three different price levels, P_3, P_2, and P_1. A decline in *P* shifts the *IS* curve to the right via the Pigou effect, the income redistribution effect, and the other effects discussed in the preceding section. The *IS* curve also shifts to the right at an unchanged price level as a result of an increase in the investment function—business persons invest more at each interest rate—or a decrease in the saving function—income recipients save less at each level of income. Lastly the *IS* curve shifts to the right as a result of expansionary fiscal actions such as a decrease in tax receipts or an increase in government spending. There is some specific fiscal action that will shift the IS_3 curve to the broken-line IS'_3 curve and will similarly shift the IS_2 curve to IS'_2 and the IS_1 curve to IS'_1. As with the *LM* curves, there are two sets of three *IS* curves in the illustration.

Suppose now that M_s happens to be the amount that gives us the solid-line set of LM curves. Combining this set with the solid-line set of *IS* curves produces the intersections *A*, *B*, and *C* that yield the curve labeled *AD* in Part B. The equilibrium *P* and *Y* are given by the intersection of this curve with the *AS* curve. What this intersection shows is a Keynesian less-than-full-employment equilibrium with actual output of Y_1 being less than Y_f. If instead of the rigid money wage rate which gives us the *AS* curve as shown, the money wage rate were perfectly flexible downward, the *AS* curve would be perfectly inelastic at Y_f and Y_1 would not be an equilibrium position as it is in the figure as drawn. According to the self-correcting mechanism of classical theory, a decline in *P* would occur to produce an intersection of *AD* with the perfectly inelastic *AS* at Y_f or equilibrium at full employment would be automatically achieved via deflation. However, it is this classical mechanism that was called into question by Keynes and others and has led us here to look into the result that follows with a rigid money wage rate and therefore with an *AS* curve like that in Part B of Figure 17-8.

With the *AS* curve as there given, there is the less-than-full-employment equilibrium at Y_1 already noted. In Keynesian theory, the solution to unemployment is an appropriate rightward shift in the *AD* curve. But in Keynesian theory there is nothing in the system to automatically bring about either the increase in investment spending or the increase in consumption spending or the combination of the two that will produce the required rightward shift in the *AD* curve. To accomplish this calls for an expansionary policy by government. Assuming the economy is not in a liquidity trap, one such policy is monetary, specifically an increase in M_s, which results in the broken-line or *LM′* set of curves in Part A. With no simultaneous governmental action taken to shift the *IS* curve, the intersections between the set of *LM′* curves and the set of *IS* curves at *D*, *E*, and *F* yield the *AD′* curve in Part B. The increase in M_s has raised the *AD* curve by just the amount that provides an intersection between that curve and the *AS* curve at Y_f. There is now an equilibrium with full employment and a higher price level. To reach the full-employment position, note that the price level must rise under present assumptions. The



A



B

FIGURE 17-8
Monetary–fiscal policies and the full-employment equilibrium

increase in employment that raises Y from Y_1 to Y_f will only occur with a decline in the real wage rate and such a decline can only occur with a rise in the price level as long as the money wage rate is inflexible downward as is assumed.

Return now to the original less-than-full-employment equilibrium given by the intersection of the *AD* and *AS* curves. A second policy to lift output to the full employment level is fiscal, specifically, an increase in government spending or a reduction in tax receipts or a combination of both which results in the broken-line or *IS'* set of curves in Part A. Assuming a balanced budget before these fiscal actions are taken, a deficit results. Assuming also that the deficit so created is not financed at all by an increase in the nominal money supply, there is no shift in the *LM* curves, and the relevant set is the original solid-line set. The intersections between these sets of *IS'* and *LM* curves occur at *G*, *H*, and *J*, and the *AD* curve derived from these intersections is *AD'* in Part B of Figure 17-8. This curve intersects *AS* at Y_f. As before, there is equilibrium with full employment.

It is apparent that Part A of the figure has been drawn so that the shift in the *IS* curves from the *IS* set to the *IS'* set and the shift in the *LM* curves from the *LM* set to the *LM'* set produce the same rightward shift in the *AD* curve. Of course, the figure says nothing about the relative magnitudes of the money supply increase, on the one hand, or the government expenditures increase or tax decrease, on the other hand, that is required for this purpose, but there is some amount of monetary expansion by itself and some amount of fiscal expansion by itself that will produce the indicated shift in the *AD* curve. An appropriate mixture of the two composed of less monetary expansion and less fiscal expansion will also produce the same result. Less monetary stimulus would reduce the extent of the rightward shift of the *LM'* curves from the position of the original *LM* curves, and less fiscal stimulus would reduce the extent of the shift of the *IS'* curves from the position of the original *IS* curves. As long as the shifts are such that the LM_2' curve and the IS_2' curve intersect at Y of Y_f, the *AD'* curve that is derived will intersect the given *AS* curve at Y_f and P_2 as now shown and therefore yield a full employment equilibrium with the minimum necessary increase in the price level. As a final word here, it may also be noted that a mixture of monetary and fiscal policies which shifts the *LM* curves rightward to the positions shown by the *LM'* curves and *also* shifts the *IS* curves rightward to the positions shown by the *IS'* curves will produce a shift in the *AD* curve from *AD* to *AD''*. This would clearly be more expansion than is needed to achieve full employment; the excess would be entirely absorbed in pulling the price level above the level needed to reach the full-employment output level.

Just as Keynes and others criticized the classical argument that a system of flexible wages and prices would automatically correct lapses from full employment, some economists have for years criticized the argument that monetary and fiscal policies can be used in a way to restore an economy suffering unemployment to full employment. We will look into some of the questions raised on the appropriateness and efficacy of monetary and fiscal policies and into other aspects in later chapters. The only purpose of these few pages has been to trace the bare mechanics of the process by which monetary and fiscal policies can shift the *IS* and *LM* curves and therefore the *AD* curve in a model in which the position of the *IS* and *LM* curves depends on the price level.

ports and exports to its overall balance of payments. Equilibrium in the balance-of-payments or external equilibrium may then be combined with the model of the preceding chapters, which deals with internal equilibrium. External disequilibrium—a balance-of-payments deficit or surplus—may cause a decrease or an increase in a nation's money supply, and the way in which this occurs is next considered. There are automatic forces that come into play to correct an external disequilibrium, but these forces differ according to whether countries are operating under a system of flexible or fixed foreign exchange rates. Our next step is to outline the adjustment process under both the flexible and fixed exchange-rate systems. At this point, the adjustment process is viewed narrowly as one of merely achieving a state of internal and external equilibrium without inquiring as to the particular position at which the internal equilibrium occurs. Because that may be at the position of full-employment output or below that output, the last section of the chapter sketches the way in which these equilibria may be realized with the internal equilibrium specifically at the full-employment level of output.

THE IS-LM MODEL INCLUDING IMPORTS AND EXPORTS

Just as the *IS* curve for the two-sector economy and the three-sector economy, respectively, show all combinations of r and Y at which $S = I$ and $S + T = I + G$, the *IS* curve for the four-sector economy shows the combinations of r and Y at which $S + T + M = I + G + X$. If $S + T + M = I + G + X$, the leakages from the income stream are equal to the injections into the income stream and the quantity of goods demanded remains unchanged. However, as for the two- and three-sector economies, the *IS* curve for the four-sector economy will show a different level of Y at which $S + T + M = I + G + X$ for each different interest rate, assuming, of course, that no part of the *IS* curve is perfectly inelastic. For a given *IS* curve, the level of Y identified by a particular interest rate is the specific amount of goods that will be demanded by the four sectors combined at that interest rate. If the economy is operating at that level of Y, that is then the actual amount of goods being supplied for the time period, or the amount of goods demanded is equal to the amount supplied, and the goods market clears. At a lower (higher) interest rate, the quantity of goods demanded is larger (smaller) and the goods market will correspondingly clear at a higher (lower) level of Y.

Derivation of the IS Function

Figure 18-1 shows the derivation of the *IS* curve for both a three-sector and a four-sector economy in order to demonstrate the change that occurs with the introduction of the fourth sector. The derivation for the three-sector case was illustrated in Figure 16-3 and discussed on pp. 287–89. In Figure 18-1, the IS_c curve for the three-sector or closed (c) economy is derived from the $I + G$ curve in Part A and the $S + T$ curve in Part C in the way earlier explained. To bring in the export side of foreign trade, the amount of X is added horizontally to the $I + G$ curve in Part A to produce the $I + G + X$ curve. The amount of X is assumed to be determined entirely by factors outside the domestic economy, and therefore the interest rate in the domestic economy does not affect exports. The $I + G + X$ curve is accordingly equidistant from the $I + G$ curve at each interest rate. To bring in the import side of foreign trade,

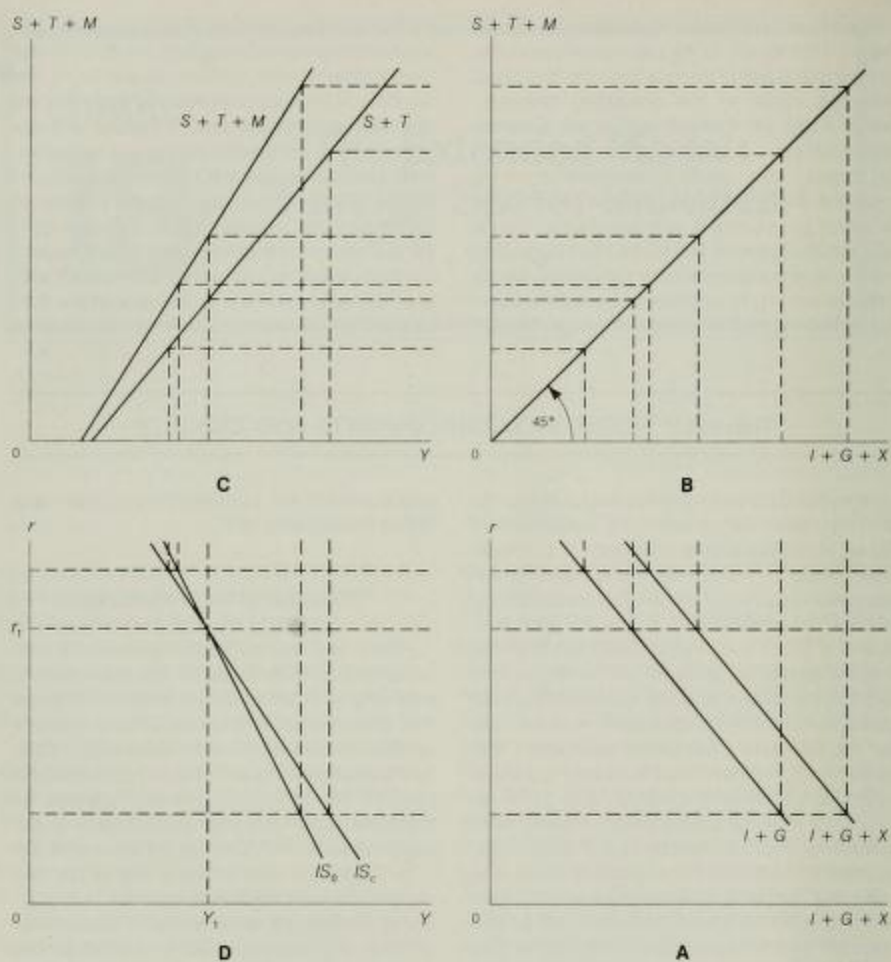


FIGURE 18-1
Derivation of the IS function

M is added vertically to the $S + T$ curve in Part C. In the model of Chapter 7 carried over here, $M = M_a + mY$ so that the spread between $S + T$ and $S + T + M$ varies directly with the level of Y. The difference between the slope of $S + T$ and the slope of $S + T + M$ is equal to m or the marginal propensity to import. The various combinations of r and Y at which the

sum of the leakages, $S + T + M$, equals the sum of the injections, $I + G + X$, are identified in the usual way. The points in the Y, r space corresponding to these combinations are connected with a line that is the IS_o curve for the open (o) economy.

The IS_o curve slopes downward to the right as does the IS_c curve, but the slope of the IS_o curve is seen to be greater than that of the IS_c curve. A decline in the interest rate raises investment spending. The increase in income that results will be that increase at which the sum of leakages from the income stream grow by an amount equal to the increase in injections, i.e., equal to the increase in investment spending. With the leakage into imports now added to the leakage into saving, the indicated increase in the sum of leakages will occur with a smaller increase in income than would be the case if there were only the leakage into saving.[2]

The relationship between the curves in Parts A and C of Figure 18-1 are such that the IS_o and IS_c curves intersect. If the fixed spread between the $I + G$ and the $I + G + X$ curves in Part A were greater than the maximum spread between $S + T$ and $S + T + M$ in Part C over the range of income,

Shifts in the IS-LM Functions

The IS curve for the open economy will shift as a result of any change in autonomous spending, whether that change originates with any of the domestic sectors or the foreign sector. To focus on the foreign sector, assume that the $I + G$ and $S + T$ curves remain constant. Then a rightward shift in the IS curve will result from either an increase in X or a decrease in M_a in the equation $M = M_a + mY$, and a leftward shift in the IS curve will result from either of the opposite changes. X is completely autonomous so the subscript a has not been attached to it. The amount by which the IS curve will shift in the present model is equal to the change in X or in M_a times $1/(1 - c + m)$, the foreign trade multiplier developed in Chapter 7.

Because the extended model with which we are now working does not assume the stable price level of the earlier model, we may also examine here the effect of a change in the domestic price level relative to foreign price levels on the IS curve. Assuming no offsetting change in the foreign exchange rates between currencies, a relative decline in the domestic price

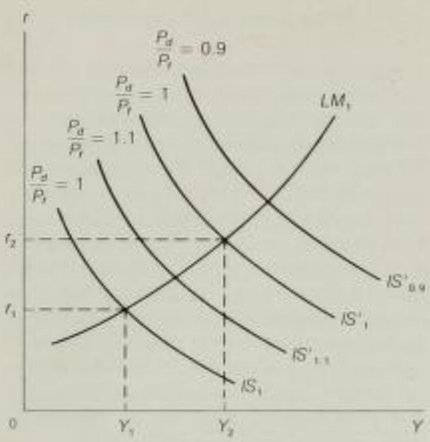


FIGURE 18-2
Shifts in the IS function

downward. A relative increase in the domestic price level will naturally have the opposite effect.

Figure 18-2 shows some of the effects of a shift in the *IS* curve brought about by foreign trade. IS_1 and LM_1 are the original positions of the curves. The intersection identifies Y_1 and r_1. Assume now a rightward shift of the *IS* curve that results from an increase in *X*, the cause of which might be an increase in incomes in other countries, a decrease in import duties in other countries, or a change in tastes of people in other countries in favor of foreign goods and foreign travel. Or the shift in the *IS* curve could result from a decrease in M_a, the cause of which might be an increase in import duties in the domestic economy or a change in tastes by people in the domestic economy away from foreign goods and foreign travel. Suppose one or more such changes shifts the IS_1 curve to IS'_1. On the assumption of no change in relative price levels or other things, the new equilibrium is given by the intersection of IS'_1 and LM_1 at which $Y = Y_2$ and $r = r_2$. This does no more than introduce another variation of the basic analysis of Chapter 16 in which we traced the same kind of change in the Y, r equilibrium as a result of a shift in the investment curve, or an increase in government purchases, or a decrease in tax receipts, each of which shifts the *IS* curve to the right.

We may turn now to the effects of a change in relative price levels or the ratio of the domestic price level, P_d, to foreign price levels, P_f. Assume that P_d/P_f is initially 1 and that this ratio is one of the factors that initially establishes the position of the *IS* curve at IS_1. At the same time that the shift of *IS* from IS_1 to IS'_1 occurs due to factors like those noted just above, assume that the relative price ratio changes from 1 to 1.1, or domestic goods become relatively more expensive, or foreign goods become relatively less expensive. Taken by itself, this change in P_d/P_f will ordinarily increase real imports and reduce real exports and thus shift the *IS* curve leftward from IS_1.[4] However, combining this change with the change that shifts *IS* from IS_1 to IS'_1 and assuming that the strength of the change in the price level ratio is less than the strength of the other change, the result will on balance be a rightward shift of the *IS*

[4]Empirical studies suggest that this is the usual result and it is the result we will accept here, but it is theoretically possible that the relative rise in the price level of domestic goods could lead to an increase in net exports or a decrease in net imports and thus bring about a rightward shift in the *IS* curve. The relatively higher cost of domestic goods will reduce the *quantity* of goods exported, but the amount of the reduction depends on the elasticity of demand for exports. At the same time, the relatively lower cost of foreign goods will increase the *quantity* of goods imported, but the amount of the increase again depends on the elasticity of the demand for imports. If the demand for exports were sufficiently inelastic and the demand for imports sufficiently elastic, the money value of imports could decline more than the money value of exports, with the result being a larger net export balance or a smaller net import balance than existed before the relative rise in the domestic price level. However, the empirical evidence does not indicate this kind of result. One study estimated the elasticity of demand for U.S. imports at 0.9 and the elasticity of demand for U.S. exports at 1.5, a combination of elasticities that will clearly result in a decrease in the value of U.S. net exports with a relative rise in the U.S. price level. See H. S. Houthakker and S. P. Magee, "Income and Price Elasticities in World Trade," *Review of Economics and Statistics*, May 1969, pp. 111–25.

curve to a position between the original position and IS'_1, say to the position shown by $IS'_{1.1}$. In the same way, if the change in the relative price ratio were from 1 to 0.9, domestic goods would become relatively less expensive and foreign goods relatively more expensive. This change, taken by itself, would shift the IS curve rightward from IS_1. In combination with the shift in the IS curve to IS'_1 due to the other change, the total shift will be to a position to the right of IS'_1, say to the position shown by $IS'_{0.9}$.

In order to identify the equilibrium Y and r that result from changes in relative price levels, we must also look at what happens to the absolute price level at the same time. Changes in the absolute price level, assuming an unchanged nominal money supply, also cause shifts in the IS and LM curves. The IS curve will shift due to a change in relative price levels as described above, but it will also shift due to a change in the absolute price level for reasons described in the preceding chapter. Relative price levels are what matters for imports and exports or foreign trade effects, but the absolute domestic price level is what matters for the Pigou, income redistribution, and tax and transfer-payment effects. However, while the IS curve will shift due to both a relative and an absolute price level change, the LM curve will only shift due to an absolute price-level change, again on the assumption of a given nominal money supply. Because a change in relative price levels is consistent with an increase, decrease, or no change in the domestic price level, various effects on the Y,r combination may follow. For example, a decrease in the P_d/P_f ratio from 1.0 to 0.9 will in itself tend to shift the IS curve rightward. If the decrease occurs in part or in whole via a decline in the domestic price level, there is an increase in the real money supply on the assumption of an unchanged nominal money supply. Thus, the rightward shift in the IS curve due to foreign trade effects will be reinforced by a rightward shift due to the Pigou and other effects that depend on the absolute price level. Also in this case the LM curve will shift rightward and the IS-LM intersection must occur at a higher Y than otherwise. On the other hand, if the decline in the ratio from 1.0 to 0.9 is entirely the result of a smaller absolute increase in the domestic price level than in foreign price levels, there is a decrease in the real money supply, again assuming an unchanged money supply. The rightward shift in the IS curve due to the foreign trade effect will be accompanied by a leftward shift due to the Pigou and other effects that depend on the absolute price level, now higher than before. Whether on balance the IS curve shifts rightward or leftward depends, of course, on the relative strengths of these effects. The LM curve in this case will clearly shift leftward. The new Y and r may be higher or lower than the original values, depending on the particular shifts in the IS and LM curves that occur.

Just as we have so far used the IS-LM apparatus to trace the effects on Y and r of changes in variables like autonomous exports, autonomous imports, and relative and absolute price levels, we can use it in the same way to trace the effects on Y and r of changes in yet other international variables, e.g., foreign exchange rates. However, while much can be seen with the IS-LM apparatus alone, much more can be seen by supplementing its IS and LM functions with a third function, a balance-of-payments or BP function. The IS-LM-BP apparatus is much more illuminating, and it is to the development and application of this broader apparatus that we turn next.

THE BALANCE OF PAYMENTS FUNCTION

The only international transactions taken into account so far have been the purchases and sales of goods and services between firms and persons in different countries or, in a word,

imports and exports. Included here are goods such as corn and computers and services such as tourism and transportation. In addition to these commercial transactions are private and public transfers. These include remittances by people in one country to relatives and friends in other countries, government transfers to foreign governments of both military goods and money and government transfers to persons abroad of money in payment of pensions. Like goods and services, transfers flow in both directions, and a net figure for transfers is reached in the same way that a net figure for exports and imports is reached. The sum of the figures for net exports of goods and services *and* net transfer payments to foreigners is, in balance-of-payments terminology, called the *balance on current account.*[5]

Beyond the transactions in goods and services and the private and government transfers which are combined to arrive at the balance on current account is a second major set of transactions. People and firms in one country spend abroad not only for goods to import into the home country and for services provided to people and firms of the home country by foreign firms, but to acquire such things as land, dwellings, plant and equipment in foreign countries, and stocks and bonds issued by foreign corporations and foreign governments. The purchases of real assets physically located abroad or financial assets issued by foreign governments and firms (including bank accounts issued by foreign firms in the banking business) are described as capital as opposed to current transactions. The balance for any time period—the difference between such purchases by persons and firms of one country in all other countries and the purchases by persons and firms of all other countries in the first country—is called the *balance on capital account.*[6]

The Balance of Payments Deficit or Surplus

During any year, each country will have a deficit or a surplus in its current account and a deficit or a surplus in its capital account. (A balance in either or both of these accounts for a year is, of course, possible, but not common.) The sum of these two figures is called the *deficit or surplus in the balance of payments* for that period. If there is a surplus in one account and a deficit in the other and if they are of the same absolute size, the sum of the two is, of course, zero. Such a zero sum is what is required for *equilibrium in the balance of payments,* although economists differ as to the number of years within which that zero sum should be achieved in order to warrant the conclusion that there is equilibrium in the country's balance of payments. If we simplify by omitting transfers, the surplus on current account is $X - M$. If we designate the deficit on capital account or the net capital outflow by H, we have the following equation for B, the balance of payments surplus: $B = (X - M) - H$. If $X - M$ equals H, the sum or B is zero and the balance of payments is in equilibrium.

Thus, if during a year the U.S. were to show exports of goods and services of \$105 billion and imports of \$100 billion, the surplus on current account would equal \$105 − \$100 or +\$5 billion. If during the year U.S. holdings of assets in foreign countries were to increase by \$30 billion and foreign holdings of assets in the U.S. were to increase by \$25 billion, there would be

[5]In the GNP identity of Chapter 2, $C + S + T + R_{pf} = GNP = C + I + G + (X - M)$, where $X - M$ is the balance on goods and services and R_{pf} is net transfer payments made abroad by private parties. The remaining items in the balance on current account, government transfer payments abroad, do not appear explicitly in the basic GNP identity. The way in which these items fit into the identity is described in Section III of the national income accounting appendix at the end of the book.

[6]A nation's GNP is a measure of currently produced goods and services and not of assets or changes in ownership of assets. The international transactions that directly enter the GNP are found in the current account; the international transactions covered in the capital account entail purchases and sales of assets and as such do not directly affect the GNP total or its separate components.

a net capital outflow of \$5 billion, or the deficit on capital account would be \$5 billion. This set of figures would indicate equilibrium in the balance of payments or $B = 0$, as (\$105 − \$100) − (\$30 − \$25) = 0.

Assuming in this illustration that all transactions are settled in foreign currencies, the U.S. would have received \$5 billion more in foreign currencies through its exports than it would have paid in foreign currencies for its imports, so it would show a gain in holdings of foreign currencies of \$5 billion for the period as a result of its transactions on current account. However, its purchases of assets in foreign countries call for the payment to foreigners of \$5 billion more in foreign currencies than the amount of foreign currencies received from foreigners for their purchases of assets in the U.S., so the U.S. would show a decline in holdings of foreign currencies of \$5 billion as a result of its transactions on capital account. In sum, the balances in the current and capital accounts offset each other and U.S. holdings of foreign currencies are unchanged. Looked at in one way, a nation's balance of payments is in equilibrium for a time period if there is neither an accumulation nor decumulation of that nation's holdings of foreign currencies during the time period. This conveys in a brief but rough way the meaning of the concept of balance-of-payments equilibrium.

The illustration above shows an equilibrium in which a current-account surplus is matched by a capital-account deficit. An equilibrium may also result with the opposite pairing. Reversing the same set of numbers, we have $X - M$ = \$100 − \$105 = −\$5 billion balance on current account and H = \$25 − \$30 (capital outflow minus capital inflow) or a balance on capital account of −\$5 billion. With $X - M$ = −\$5 billion and H = −\$5 billion, the sum of $(X - M) - H$ is zero, and there is again a balance-of-payments equilibrium.

Of course, various changes in the illustrative numbers will convert the equi

balance-of-payments equilibrium, any reduction in the net export balance or increase in the net import balance will result in a balance-of-payments deficit, assuming no change in the balance on capital account. And any increase in the net export balance or decrease in the net import balance will result in a balance-of-payments surplus on the same assumption. Starting again from a balance-of-payments equilibrium, any increase in the net capital outflow or decrease in the net capital inflow will result in a balance-of-payments deficit, assuming no change in the balance on current account. And any decrease in the net capital outflow or increase in the net capital inflow will result in a balance-of-payments surplus on the same assumption.

Derivation of the Balance-of-Payments Function

Given the simplification here adopted that the balance on current account is equal to the net import or net export balance, the determination of the balance on current account may be explained with the model of Chapter 7. That simple model makes exports entirely a function of factors outside the domestic economy and makes imports a function of the income level within the domestic economy. In line with this model, Part C of Figure 18-3 shows that the $X - M$ balance varies inversely with the domestic income level. The higher Y, the smaller $X - M$ or the smaller the net export balance or the larger the net import balance.

What determines the balance on capital account? People and firms purchase those assets that provide the highest yield, risk taken into account, and they will switch from assets in their own countries to assets in other countries if the difference in yield, adjusted for differences in risk, makes such a switch advantageous. Therefore, the net capital outflow, H, may be written as an inverse function of the rate of

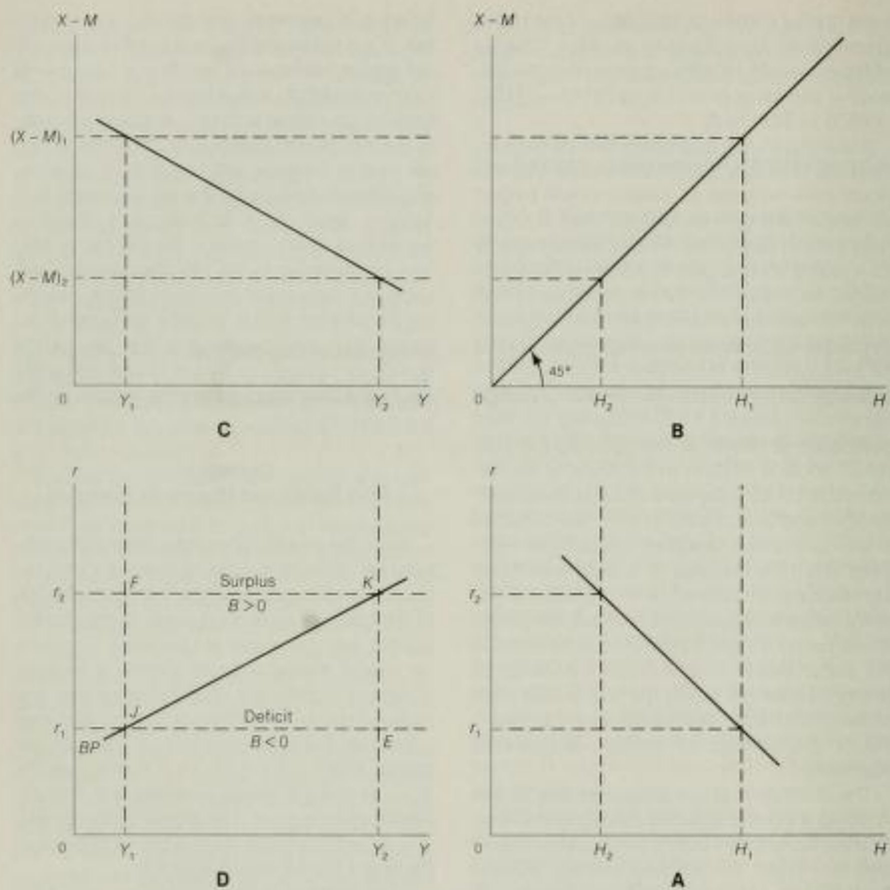


FIGURE 18-3
Derivation of the BP function

the lower the rate in the domestic economy, the larger is the net capital outflow or the smaller is the net capital inflow. The curve in Part A of Figure 18-3 shows this relationship between H and r.

Part B of the figure provides a 45° line needed to tie together H in Part A and $X - M$ in Part C. Each point along the 45° line shows an equality between the value of $X - M$ on the vertical axis and H on the horizontal axis and thus a zero value for B.

In the same way that the *IS* function was derived in Part D of Figure 16-3, p. 228, the *BP* function in Part D of Figure 18-3 is derived

from the other parts of this figure. Suppose that the interest rate in Part A is r_1. This indicates a capital outflow of H_1. To secure B of zero needed for equilibrium in the balance of payments, $X - M$ must equal H. Tracing from Part A through Part B to Part C indicates that $X - M$ will be equal to H_1 if Y is equal to Y_1. Bringing together in Part D the values of r_1 from Part A and Y_1 from Part C gives us one combination of r and Y at which $B = 0$. To vary the approach, start this time in Part C to derive a second combination. Suppose the income level is Y_2, which indicates $(X - M)_2$. Balance-of-payments equilibrium requires that H equal $(X - M)_2$. Tracing from Part C through part B to Part A indicates that H will be equal to $(X - M)_2$ if r is r_2. Bringing together in Part D the values of r_2 and Y_2 gives us a second combination of r and Y at which $B = 0$. Connecting the points identified by these and other such combinations yields the BP function.

The BP function will always slope upward to the right or balance-of-payments equilibrium will be found at combinations of relatively high interest rates and relatively high income levels or relatively low interest rates and relatively low income levels. This follows from our assumptions that a relatively high income level will mean a relatively low net export surplus; equilibrium in the balance of payments then requires a relatively low net capital outflow, which will be secured with a relatively high domestic interest rate. Or starting with a relatively low domestic interest rate, there will be a relatively large net capital outflow; equilibrium in the balance of payments then requires a relatively large net export balance, which will occur with a relatively low level of income.

Every point along the BP curve identifies a combination of r and Y at which $X - M = H$ or $B = 0$. Every combination not on the curve is one at which $X - M \gtreqless H$ or $B \gtreqless 0$. Specifically, all combinations above the BP line are combinations at which $X - M > H$ and therefore combinations at which there is a balance-of-payments surplus; all combinations below the BP line are combinations at which $X - M < H$ and therefore combinations at which there is a balance-of-payments deficit.

This may be seen as follows. The combination of Y_2, r_2 shown in Part D of Figure 18-3 by K, for example, is one at which there is equilibrium. As the combination of Y_1, r_2 shown by F includes the same interest rate as at K but a lower income level than at K, H will be the same as at K but $X - M$ will be larger than at K. Therefore, because $X - M = H$ at K, $X - M$ must exceed H at F or there must be a balance-of-payments surplus at F. The same may be seen by comparing the combination at the point F with that at J. The combination of Y_1, r_1 shown by J is one at which there is equilibrium. As the combination of Y_1, r_2 shown by F includes the same income level as at J but a higher interest rate than at J, $X - M$ at F will be the same as at J but H will be smaller at F than at J. Since $X - M = H$ at J, $X - M$ must exceed H at F or there must be a balance-of-payments surplus at F. The same conclusion—a balance-of payments surplus—will be found for any other combination of Y and r that lies above the BP curve.

By the same reasoning, a comparison of the Y, r combination shown by E, for example, with the equilibrium combinations shown by J and K will reveal that there must be a balance-of-payments deficit for the Y, r combination given by E. The same will be found for any other combination of Y and r that lies below the BP curve.

Shifts in the Balance-of-Payments Function

If the actual interest rate and income level in the economy are at a point like E or F or at any other point not on the BP function, the deficit or surplus in the balance of payments tends to automatically bring about an equilibrating shift in the BP function. Thus, in the case of a surplus, the BP function in Part D of Figure 18-3 will shift upward or leftward and in the case of a deficit downward or rightward. We conclude

this section by looking at the way a change in the price level and a change in the foreign exchange rate will shift the BP function. Here we merely see why such changes will shift the BP function. Later we will look at the automatic forces in the system that tend to bring about changes in the price level and changes in the foreign exchange rate.

A Change in the Price Level Part C of Figure 18-4 shows an initial net export curve labeled $(X - M)_1$ in which the subscript 1 identifies P of P_1. What is the effect of an absolute change in the domestic price level on the *BP* function, other things being equal? We saw above that an absolute rise in the domestic price level decreases the net export balance and shifts the *IS* curve leftward and also decreases the real money supply and shifts the *LM* curve leftward. In Figure 18-4 we find that an absolute rise in the domestic price level from P_1 to P_2 shifts the $(X - M)$ curve leftward from $(X - M)_1$ to $(X - M)_2$ and this, in turn, shifts the *BP* curve in Part D leftward from BP_1 to BP_2. The result is that equilibrium in the balance of payments at any level of Y now requires a higher r or, put another way, equilibrium at any r now requires a lower Y. As the rise in P from P_1 to P_2 reduces the level of X at each level of Y and increases the level of M at each level of Y, equilibrium in the balance of payments requires either that Y fall enough to restore $X - M$ to what it was before the rise in P, namely, a decrease of FJ in Part D, or that r rise enough to reduce H by an equal amount, namely, an increase of FK in Part D, or any combination of the two. A decline in the domestic price level from P_1 to P_0 will naturally have the opposite effect as shown by the rightward shift from $(X - M)_1$ to $(X - M)_0$ in Part C and the resulting rightward shift from BP_1 to BP_0 in Part D.

A Change in the Foreign Exchange Rate A foreign exchange rate is simply the price of one currency in terms of another or the quantity of one currency needed to purchase a unit of another. For example, in the last few years, the price of the West German currency unit, the Deutsche mark or DM, has been around $0.40 in terms of the U.S. currency unit, the dollar, or, expressed the other way, the price of the U.S. dollar has been around 2.5 DMs. For this pair of currencies, an appreciation or rise in the dollar means that the buyer of dollars must pay more than 2.5 DMs per dollar. Or looked at the other way, an appreciation or rise in the dollar, which is a depreciation or fall in the DM, means that the buyer of DMs need pay less than $0.40 per DM. To prospective West German buyers of U.S. goods and services, the attractiveness of those goods depends, among other things, on their prices in terms of dollars and on the price of the dollar in terms of the DM. Basically, a 10-percent rise in dollar prices of U.S. goods with no change in the price of the U.S. dollar in terms of the DM is the same to the prospective German buyer as a 10-percent rise in the price of the U.S. dollar in terms of the DM and no change in the prices of U.S. goods in terms of dollars. For example, if a U.S.-manufactured computer is priced at $100,000 and the U.S. dollar is priced at 2.5 DMs, the computer will cost a German importer 250,000 DMs. To that importer, a 10 percent rise in the price of the computer to $110,000 with no change in the price of the dollar ($1 = 2.5 DMs) will raise the cost to 275,000 DMs and a 10 percent rise in the price of the dollar ($1 = 2.75 DMs) with no change in the price of the computer will also raise the cost to 275,000 DMs.

If we now assume a rise in the foreign exchange rate of the dollar against other currencies in general, i.e., that it takes more DMs, pesos, francs, lira, escudos and so forth than before to buy a U.S. dollar in the foreign exchange markets, it will be seen from the argument just presented that such a rise in the foreign exchange rate will cause the *BP* curve to shift leftward just as a rise in the U.S. price level will cause the *BP* curve to shift leftward. Assuming here that there is no change in the price levels of the other countries, U.S. ex-

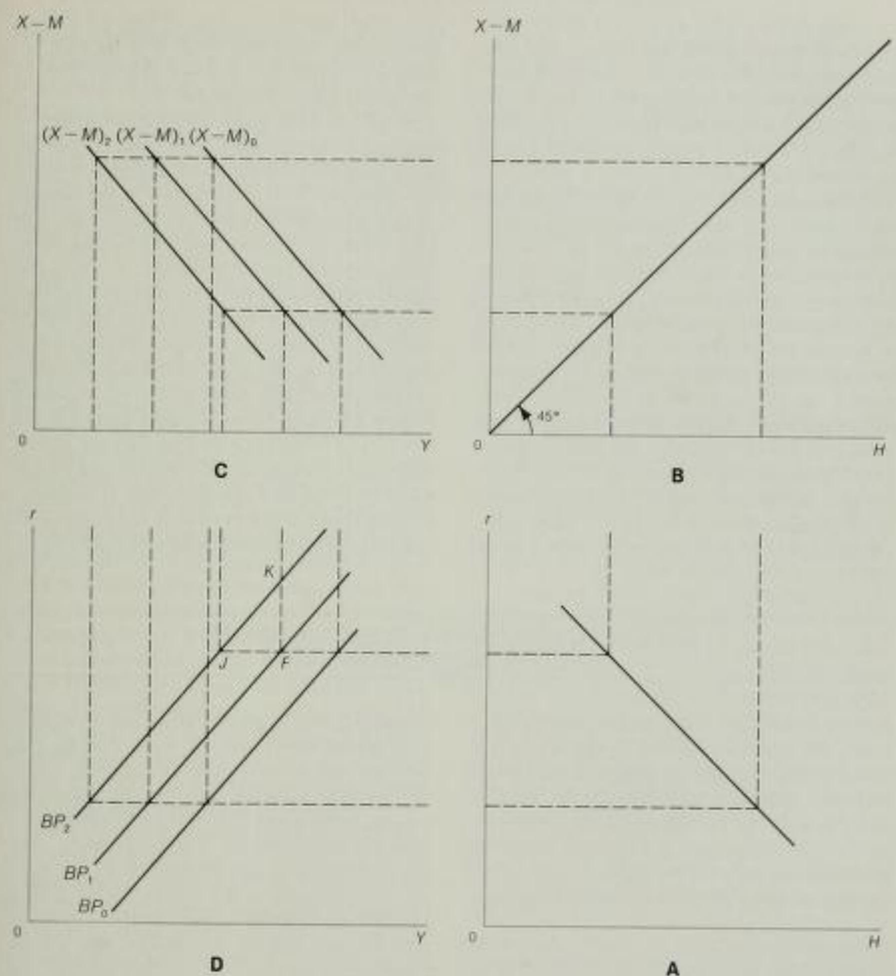


FIGURE 18-4
Shifts in the BP function

ports will decline because it now takes more of each foreign currency unit to purchase the U.S. currency unit, and U.S. imports will grow because the U.S. currency unit will purchase more of each foreign currency unit. In Part C of Figure 18-4, a rise in the foreign exchange rate—a rise of the dollar in terms of other currencies—will shift the $X - M$ curve from

$(X - M)_1$ to $(X - M)_2$, the same kind of movement that occurs with a rise in the price level. This leftward shift in the $X - M$ function, in turn, shifts the BP function leftward from BP_1 to BP_2, with the result that balance-of-payments equilibrium at any Y requires a higher r or at any r requires a lower Y.[7]

By the same argument, a fall in the foreign exchange rate of the dollar against other currencies in general will cause the BP curve to shift rightward just as a fall in the U.S. price level will cause the BP curve to shift rightward. As foreigners can now purchase a U.S. currency unit with less of their own units, they will buy more U.S. goods or U.S. exports will increase; as U.S. persons and firms can now purchase fewer foreign currency units with each dollar, they will buy less foreign goods and U.S. imports will decrease. For this case, the $X - M$ curve in Part C of Figure 18-4 will shift rightward from $(X - M)_1$ to $(X - M)_0$, the same kind of shift that occurs with a lower price level. The BP function, in turn, shifts from BP_1 to BP_0, with the result that balance-of-payments equilibrium at any Y requires a lower r or at any r requires a higher Y.

What has here been said for the effect of changes in foreign exchange rates on the BP function also applies to the IS function. In terms of Figure 18-1, p. 332, a decline in the foreign exchange value of a nation's currency will shift the $I + G + X$ curve rightward in Part A by increasing X and will shift the $S + T + M$ curve in Part C downward by decreasing M_a. Each of these shifts tends to shift the IS_0 curve in Part D rightward. A rise in the foreign exchange value of a nation's currency will by the same reasoning bring about a leftward shift in the IS_0 curve in Part D. Thus, when we later combine the IS, LM, and BP functions in one diagram, we will note that changes in the foreign exchange value of a nation's currency, like changes in its price level relative to foreign price levels, affect the position of both the BP and IS functions.

BALANCE-OF-PAYMENTS DISEQUILIBRIUM AND THE MONEY SUPPLY

If there is equilibrium in the balance of payments, the total payments to foreigners for the sum of goods and services imported and for real and financial assets acquired abroad are just equal to the total receipts from foreigners for their imports of goods and services from the domestic economy and for their real and financial assets acquired in the domestic economy. However, a surplus means an excess of receipts from abroad over payments made abroad, and a deficit means an excess of payments made abroad over receipts from abroad. Such an imbalance between payments and receipts or between outpayments and inpayments may cause changes in the economy's money supply, and changes in the money supply play a role in the process by which a nation achieves external and internal equilibrium. We will be turning to this process in the next section, but we first look here at the way that payments imbalances may bring about changes in a country's money supply.

Exporters and others who receive foreign currencies (which are mostly in the form of bank drafts or claims to foreign currency rather than the paper money itself) sell these cur-

[7]As explained in footnote 4, it is theoretically possible for a rise in a country's price level relative to foreign price levels to result in an increase in the money value of its net exports or a decrease in the money value of its net imports, if that country's demand for imports is sufficiently elastic and other countries' demand for its exports is sufficiently inelastic. The same is true for foreign exchange rates. It is theoretically possible for a rise in a country's foreign exchange rate to result in an increase in the money value of its net exports or a decrease in the money value of its net imports, if the price elasticities of demand are as described.

rencies to commercial banks in exchange for their own currency, and importers and others who make payment in foreign currencies buy the needed foreign currencies with their own currency from the banks. If a country has a balance of payments surplus for any given time period, the commercial banks as a group will find an increase in their overall holdings of foreign currencies over that period. Each day the amount of foreign currencies purchased is that much added to the banks' assets and the payment for these currencies, which is typically in the form of credits to the demand deposit balances of the sellers, is that much added to the banks' demand deposit liabilities. Similarly, the sale of foreign currencies each day is that much deducted from the banks' assets and the payment for these currencies, which is typically in the form of reductions in the demand deposits of the currency buyers, is that much subtracted from the banks' demand deposit liabilities. With a balance-of-payments surplus, there is a net increase in the amount of foreign currencies held by the country's banks. Other things being equal, there is then a net increase in the banks' assets and an equal net increase in the banks' demand deposit liabilities to the public. Thus, with a net increase in the public's holding of demand deposits and no change in its holdings of currency, there is an increase in the country's money supply as a result of the surplus in its balance of payments.

One may object that this conclusion is valid only if the commercial banks are willing and able to hold among their assets larger amounts of foreign currencies. What if these banks choose not to do so or are unable to do so? For illustration, take one specific currency, say the French franc, and assume that all countries on balance show an increase in holdings of francs or, in other words, that France shows a deficit in its balance of payments for the period. If U.S. commercial banks choose not to hold the additional francs they have purchased from some of their customers but not sold to other customers, how may they dispose of them? Under the international payments system which prevailed from the end of World War II until 1973, unwilling holders of francs could in effect sell them to the French central bank, the Bank of France. The U.S. commercial banks in practice would sell the francs to the Federal Reserve Banks in exchange for additional reserve balances at the Federal Reserve Banks at a fixed price of so many francs per dollar of additional reserves. The Federal Reserve Banks in turn would in effect present the francs to the Bank of France in exchange for dollars at the same fixed price. If the Bank of France did not have dollars or other currencies acceptable to the Federal Reserve Banks, it would have to make payment in gold, again at a fixed price of so many francs per ounce of gold. The international monetary system was one of fixed foreign exchange rates, and under this system the Bank of France would act in the way described—exchange its holdings of U.S. dollars or gold for francs presented to it by foreign central banks—in order to prevent a decline in the value of the franc in terms of the U.S. dollar or gold.

It may now be seen that carrying the story through this second step actually reinforces the conclusion we wish to show, namely, that a country with a surplus in its balance of payments will show an increase in its money supply, other things being equal. For at this second step there is not only the original increase in the money supply identified above, but now the potential for a further increase. The U.S. commercial banks have converted foreign currency holdings into additional reserves at the Federal Reserve Banks which increase their ability to extend credit to their regular customers and thereby further increase demand deposits and the money supply. Because each dollar of reserves can support a multiple of that amount in demand deposits, the amount of credit extended and the new money created is also a multiple of the increase in commercial bank reserves.

Consider now a deficit in a country's balance of payments. This should have the opposite

effect of a surplus, namely, a reduction in the country's money supply, assuming, of course, that other things remain equal. For a time period during which there is a deficit, the nation's banks will find that the amount of foreign currencies they sell to importers and others who make payments abroad exceed the amount they buy from exporters and others who receive payments from abroad. Because the bank customers buying foreign currencies draw down their demand deposit balances more than the bank customers selling foreign currencies build up their demand deposits, the banks show on the asset side a net decrease in holdings of foreign currencies and on the liability side a net decrease in demand deposits. Thus, with a net decrease in the public's holdings of demand deposits and no change in its holdings of domestic paper money and coin, there is a decrease in the country's money supply as a result of the deficit in its balance of payments.

While this is one way in which things may work out, it is a way that will occur only if the commercial banks have "in stock" the amount of foreign currencies needed to meet the excess of the amount importers and others wish to purchase from them over the amount exporters and others have to sell to them. If they do not have this amount, the deficit may be financed out of the foreign currency holdings of the central bank, which are maintained by the central bank to meet just such needs. The central bank could sell the needed foreign currencies to importers and others whose requirements could not be met by the commercial banks. In payment the central bank would receive checks drawn against the demand deposits of the buyers in the commercial banks. As the commercial banks maintain reserve balances with the central bank, the amount of these checks will be deducted from these balances by the central bank. We are now at the step that parallels the other case. In the process as described, the commercial banks lose reserves equal in amount to the sales of foreign currencies by the central bank to importers and others in the domestic economy. Other things being equal, the loss of reserves may force the commercial banks to contract credit by a multiple of this loss and this means a corresponding decrease in demand deposits and the money supply.

Under the international payments system that prevailed from World War II to 1973, those who sought to purchase foreign currencies with their own currency were able to do so at a fixed rate. Each country was committed to maintaining a fixed value for its currency in terms of other currencies. For a deficit country to prevent a decline in the external value of its currency, the central bank in the country had no choice but to use its reserves of foreign currencies or gold to meet the demands of those who wished to convert the domestic currency into foreign currencies. As long as the central bank had the reserves to meet such demands and was willing to use them, the external value of the nation's currency could be maintained.[8]

Although it bypasses many complications, the review here offered is sufficient to show the way in which a balance-of-payments deficit or surplus tends to decrease a deficit country's money supply and increase a surplus country's money supply under an international monetary system of fixed exchange rates. However, since 1973 the system has been one of flexible exchange rates, a system in which central banks

[8]Suppose the deficit country in the illustration is France. The Bank of France will be faced with demands from importers in France for foreign currencies, say U.S. dollars, to settle transactions in which it was agreed that payment was to be made in dollars. The central bank accordingly sells dollars to its own nationals in exchange for francs, as in the present illustration. In the earlier illustration in which commercial banks in the U.S. and other countries found themselves with enlarged holdings of francs due to the French deficit, U.S. and other exporters had accepted payment for goods sold to France in francs and had sold those francs to their commercial banks for domestic currency. In that illustration, the U.S. commercial banks exchanged these francs for dollars at the Federal Reserve Banks and the latter in turn presented the francs to the Bank of France for payment in dollars. The only difference in the two illustrations of the French deficit is that the demand for dollars from the Bank of France originates in the first illustration with U.S. exporters who accepted francs in payment and in the other illustration with French importers who have agreed to make payment in dollars.

do not stand ready to use their holdings of foreign currencies to stabilize the value of their currency at a fixed level in terms of other currencies. Under flexible exchange rates, a deficit country lets the external value of its currency depreciate and a surplus country lets the external value of its currency appreciate. If this process is permitted to run its course freely, a balance-of-payments deficit or surplus that does appear will be a short-lived phenomenon. The foreign exchange rate established in a free market by the forces of supply and demand will move to the level that eliminates a deficit or a surplus. To the degree that deficits and surpluses are corrected in fairly short order under a system of flexible exchange rates, there is no automatic tendency for a deficit or a surplus to cause a decrease or increase in the money supply in the way found under a system of fixed exchange rates.

BALANCE-OF-PAYMENTS DISEQUILIBRIUM AND THE ADJUSTMENT PROCESS

Using the groundwork laid in the preceding pages, we now turn to an examination of the processes by which a surplus or deficit in the balance of payments may be eliminated. In Part A of Figure 18-5, internal equilibrium is defined by the Y_1, r_1 combination at which the *IS* and *LM* curves intersect. However, because the Y_1, r_1 intersection lies above or to the left of the *BP* function, there is external disequilibrium, specifically a surplus in the balance of payments. In Part B of Figure 18-5, because the Y_1, r_1 intersection that provides internal equilibrium lies below or to the right of the *BP* function, there is the opposite kind of external disequilibrium, a deficit.

In terms of Figure 18-5, it will be seen that

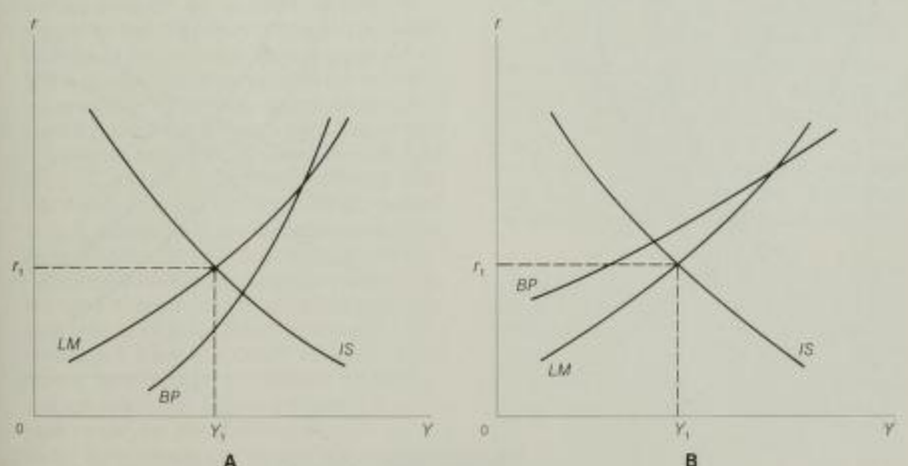


FIGURE 18-5
Balance-of-payments disequilibria

to wipe out a balance-of-payments deficit or surplus requires that the *IS-LM* intersection lie neither above nor below the *BP* function or, what is the same thing, that the *BP* function lie neither below nor above the *IS-LM* intersection; to achieve this result requires shifts in one or more of the three functions that will produce an intersection of all three functions at one *Y,r* combination. The adjustment process that brings about shifts in one or more functions varies according to whether the international monetary system is one with flexible or fixed exchange rates. During most of the period since World War II, the international economy operated under the fixed exchange-rate system established by the major powers at Bretton Woods, New Hampshire in 1944. However, the Bretton Woods system weakened during the sixties and collapsed in the early seventies. During the years since then, the world has lived with a flexible exchange-rate system or a system of floating exchange rates, but a return to a form of a fixed exchange-rate system is possible in the nineteen-eighties. The process of adjustment to a balance-of-payments disequilibrium differs under the two systems, or, in terms of Figure 18-5, the shifts in the functions differ under the two systems. The basic mechanics of the adjustment process under each of the two systems will here be traced in turn.

The Flexible Exchange-Rate System

Take first the case of a surplus indicated by the *LM*, IS_1, and BP_1 set of curves in Part A of Figure 18-6. The quantity of foreign currencies received from foreign purchasers of goods and services and foreign purchasers of real and financial assets in the domestic economy during the time period exceeds the amount of

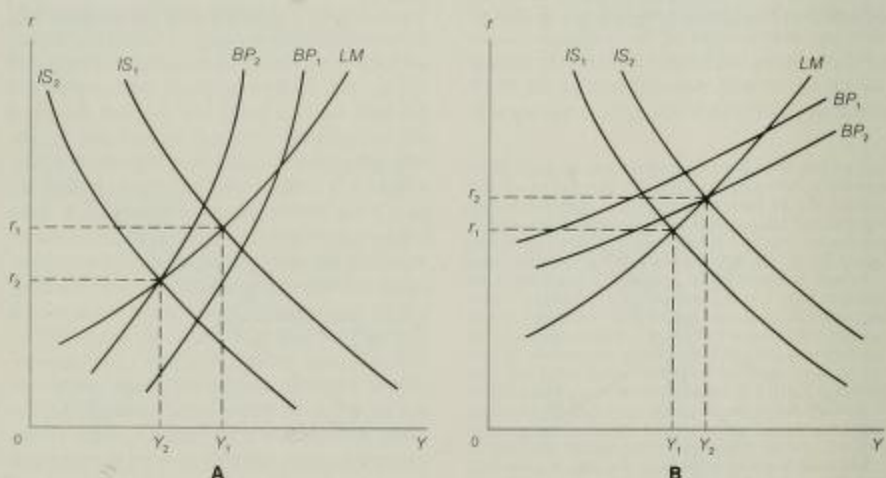


FIGURE 18-6
Balance-of-payments equilibrium under flexible exchange rates

foreign currencies needed to make payment for goods and services purchased from and assets acquired in foreign countries during that period. Under the flexible exchange-rate system, the foreign exchange value of the economy's currency is not stabilized within a narrow band by the central bank, but is essentially allowed to find its own level in response to the forces of supply and demand in the free market. In the case of a surplus, the excess supply of foreign currencies means that it will take more units of such currencies than before to exchange for a unit of the domestic economy's currency, or the foreign exchange value of the domestic economy's currency will rise. This discourages exports and encourages imports, the result of which is to produce a shift to the left of both the *BP* and *IS* functions, say, from IS_1 and BP_1 to IS_2 and BP_2 in Part A of Figure 18-6. With these two shifts, the *IS*, *LM*, and *BP* functions all intersect at Y_2, r_2 at which income level and interest rate there is both internal and external equilibrium.

Part B of Figure 18-6 shows the adjustment in the case of a deficit. In this case the amount of foreign currencies needed to make payments abroad exceeds the amount received from abroad so that the prices of foreign currencies in terms of the domestic currency are bid up or the foreign exchange value of the domestic currency falls. Accordingly, exports are encouraged and imports are discouraged. The *BP* and *IS* functions both shift rightward from IS_1 and BP_1 to IS_2 and BP_2 in Part B of Figure 18-6. After this the *IS*, *LM*, and *BP* functions all intersect at Y_2, r_2 at which income level and interest rate there is both internal and external equilibrium.

Under the flexible exchange-rate system, the economy's money supply does not automatically tend to increase if there is a surplus or to decrease if there is a deficit. However, if the central bank sought to maintain a fixed exchange rate in the face of a surplus, it would have to put into the economy more of its currency as it bought up foreign currencies to prevent the excess of foreign currencies from raising the value of its currency. If it sought to maintain a fixed exchange rate in the face of a deficit, it would have to take the opposite action, which would reduce the amount of its own currency in its economy. Because the central bank does not add to or subtract from its own money supply in operating under the flexible exchange-rate system, the *LM* curve does not automatically tend to shift as a part of the balance-of-payments adjustment process. The adjustment occurs as described through shifts in the *BP* and *IS* functions.[9]

The Fixed Exchange-Rate System

With the initial set of curves given by IS_1, LM_1, and BP_1 in Part A of Figure 18-7, there is a balance-of-payments surplus. The initial *Y,r* equilibrium is at Y_1, r_1. To prevent a rise in the foreign exchange value of its currency that otherwise will result with a surplus, the central bank finds it necessary to absorb the excess supply of foreign currencies and in the process increases the amount of its currency and the nominal money supply held by the public. This shifts the *LM* curve to the right. If the price level were to remain stable in the face of the expanding money supply, it would be possible to eliminate the surplus and achieve equilibrium solely through a shift of the *LM* curve from LM_1 to LM_3. The LM_3, IS_1, and BP_1 curves all intersect at the Y_3, r_3 combination of income level and interest rate. However, a significant increase in the price level may result from the increase in demand generated by what can be a sizable

[9]Under a fully flexible exchange-rate system, deficits or surpluses tend to be corrected in a relatively short time. Although the adjustment process calls for some shift in the *IS* curve and thus some shift in the aggregate-demand curve, the effects on the domestic price level of the shift in the aggregate-demand curve will not be so large as to make the price-level change play a major part in the adjustment process. Thus, the *LM* curve may be assumed to remain in its initial position, an assumption that would not be permissible with a significant change in the price level because that would change the real money supply and the position of the *LM* curve.

expansion of the money supply. There will then be a decrease in exports and increase in imports, which will cause the IS and BP curves to shift leftward from their original positions. With a rise in the price level as a result of a rightward shift in LM, the shift in LM needed to achieve equilibrium is less than the shift from LM_1 to LM_3. In Part A of Figure 18-7, equilibrium is achieved with whatever expansion of the money supply is required to shift LM from LM_1 to LM_2, because this results in that rise in the price level that shifts IS_1 and BP_1 leftward to IS_2 and BP_2.[10] IS_2, LM_2, and BP_2 intersect at Y_2, r_2 and equilibrium is established.

[10]The greater the effect on the price level of any increase in the nominal money supply, the greater will be the increase in the nominal money supply needed to produce a given rightward shift in the LM curve. The LM curve shifts rightward only if there is an increase in the real money supply, so a nominal increase in the money supply will shift the curve in this direction only if it more than offsets the tendency of the rise in the price level it generates to shift the LM curve in the opposite direction.

The initial set of curves, IS_1, LM_1, and BP_1, in Part B of Figure 18-8 indicates a deficit. The initial Y, r equilibrium is at Y_1, r_1. With a deficit, the foreign exchange value of the currency will tend to fall. However, to maintain a stable exchange value, which is the objective under the fixed exchange-rate system, the central bank must dip into its holdings of foreign currencies to buy up the excess supply of its own currency being offered in the market. This support policy reduces the amount of its currency or the nominal money supply held by the public. The balance of the argument then parallels that traced for the surplus case. If there were no decrease in the price level in the face of the decrease in the money supply, equilibrium could be achieved at Y_3, r_3 by a leftward shift in the LM curve from LM_1 to LM_3. However, with a decline in the price level, there is a smaller leftward shift in the LM curve and a rightward shift in both the IS and BP curves. Equilibrium occurs at Y_2, r_2 where IS_2, LM_2, and BP_2 intersect.

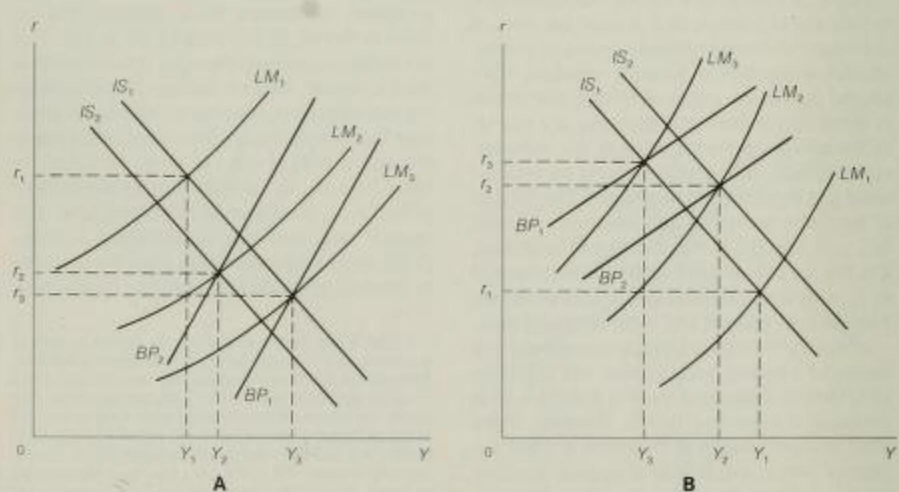


FIGURE 18-7
Balance-of-payments equilibrium under fixed exchange rates

BALANCE-OF-PAYMENTS EQUILIBRIUM AND FULL-EMPLOYMENT EQUILIBRIUM

The combination of Y and r at which the IS and LM curves intersect has been described as one consistent with internal equilibrium, but the level of Y so identified may be at or below full employment. If it is below, there is an equilibrium in the sense that an upward-sloping aggregate-supply curve is intersected by the aggregate-demand curve to the left of the output level at which

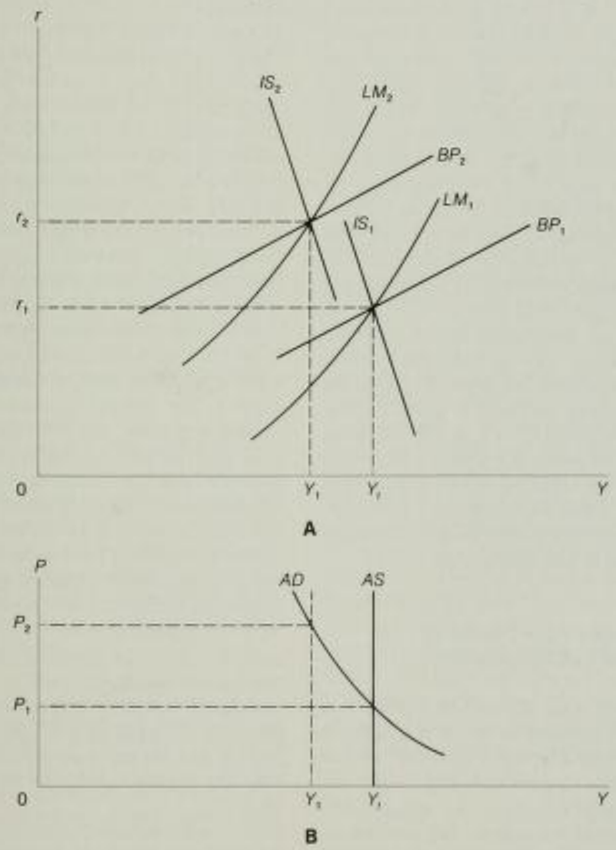


FIGURE 18-8
Balance-of-payments and full-employment equilibria through wage–price flexibility

the aggregate-supply curve becomes perfectly inelastic. That is, the *AS-AD* intersection is below the full-employment level and there is a Keynesian less-than-full-employment equilibrium. However, on the classical assumption of downwardly flexible wages and prices, the aggregate-supply curve is a vertical line at the full-employment level of output and a level of output below this is not an equilibrium level. In the preceding chapter, we looked into the process by which a deflation of wages and prices, according to the classical argument, would move the actual output level to the full-employment level as the deflation brought about rightward shifts in the *IS* and *LM* functions. Here we want to do the same thing for the model that includes a *BP* as well as *IS* and *LM* functions and therefore to show how both full-employment equilibrium and balance-of-payments equilibrium may be achieved. We also saw in the preceding chapter that with the downwardly rigid wages and prices assumed in the Keynesian model, expansionary monetary and fiscal policies are needed to shift the *IS* and *LM* functions rightward to get to the full-employment equilibrium. As a second step, here we want to see how monetary and fiscal policies may be used to achieve both full-employment equilibrium and balance-of-payments equilibrium in the model that contains a *BP* as well as *IS* and *LM* functions.

Wage–Price Flexibility and Full Employment

The IS_2, LM_2, and BP_2 curves in Part A of Figure 18-8 all intersect at Y_1 and r_2 or there is both internal and external equilibrium at this income level and interest rate. The initial positions of the curves at IS_2, LM_2, and BP_2 are based on an initial price level of P_2. As the IS_2 and LM_2 curves in Part A intersect at Y_1, Part B of the figure shows that at the price level P_2 the amount of goods demanded is Y_1 or an amount below the full-employment output of Y_f. On classical assumptions, the unemployment that exists with actual output at Y_1 leads to falling money wage rates followed by falling prices. We found in the preceding chapter that the fall in the price level would shift the *LM* curve rightward via the real-balance effect and would shift the *IS* curve rightward via the real-balance, foreign-trade, and other effects. The foreign-trade effect—an increase in exports relative to imports—will also shift the *BP* curve rightward. There is some decrease in the price level, specifically from P_2 to P_1 in the illustration, which shifts each of the three curves in such a way as to produce a new intersection of the three curves at the Y_f level of output. With a P of P_1, the IS_1 and LM_1 curves show that the amount of output demanded will be Y_f or there is equilibrium at the full-employment level of output.

The decline in the price level stimulates exports and cuts back imports and thereby increases net exports and the level of domestic output and income. However, the rise in income in turn enlarges the amount of imports, which has the opposite effect on domestic output and income. The domestic economy must be importing a greater real amount of goods at Y_f than at Y_1, given a positive marginal propensity to import, but the illustration indicates that the increase in its exports created by the decline in the price level is greater than the increase in its imports created by the rise in income. This follows from the fact that the new balance-of-payments equilibrium is at a lower interest rate and is therefore one with a larger net capital outflow. Given the equation for equilibrium in the balance of payments, $B = (X - M) - H$, because H has risen with a lower r, $X - M$ has risen by an equal amount. In absolute terms both X and M have risen with the rise in X exceeding the rise in M by the amount of the rise in H.[11]

The rise in Y from Y_1 to Y_f is due only in part to the rise in $X - M$. In a closed economy, the

[11] The adjustment process would differ in one way if we had started with an initial set of curves which, like Parts A

same decline in the price level would shift the IS and LM curves rightward and raise the level of output above Y_1. To separate out the amount of the expansion in output due to changes in real domestic spending and real foreign spending requires that one examine the curves from which the IS and LM curves are derived.

Monetary–Fiscal Policies and Full Employment

Some of the problems confronted in relying on wage–price deflation as the solution to unemployment were noted in the preceding chapter. To Keynes the problems ruled out deflation as a realistic course of action, even if the wage and price cuts that it calls for could be secured in practice. Hence, Keynes looked at the economy as it would operate with downwardly rigid wages and prices. In place of the perfectly inelastic AS curve at Y_f in the classical model of Part B of Figure 18-8, in the Keynesian model there is the upward-sloping AS curve of Part B of Figure 18-9.

Part A of Figure 18-9 shows an initial set of curves, IS_1^1, LM_1^1, and BP_1, all of which are consistent with a price level of P_1 in Part B. This set yields neither a full-employment equilibrium nor a balance-of-payments equilibrium. The intersection of the IS_1^1 and LM_1^1 curves at J indicates that Y_1 of goods are demanded at P_1 and this identifies the P_1, Y_1 point on the AD_1 curve in Part B. The AS curve shows that Y_1 of goods are supplied at P_1, so $AD = AS$ at Y_1. There is an internal equilibrium, albeit a less than full-employment equilibrium. On the external side, there is a disequilibrium, a deficit, as the intersection of IS_1^1 and LM_1^1 at J occurs below the initial BP function, BP_1.

The full-employment level of output identified in Part B is carried up as the Y_f vertical line running through Part A. It may then be seen in Part A that to meet *only* the objective of full-employment equilibrium requires that the IS and LM curves intersect anywhere along the Y_f line but to meet the objective of balance-of-payments equilibrium requires in addition that the IS, LM, and BP functions all intersect at the same point along the Y_f line.

We saw in the preceding chapter that through expansionary monetary policy (barring an inconsistency between saving and investment or a liquidity trap) government could presumably shift the LM curve rightward as needed to reach the full-employment output or through fiscal policy it could presumably shift the IS curve rightward to achieve the same objective. The same may be seen here. Through an appropriate expansion of the nominal money supply, the LM curve may be shifted from LM_1^1 to LM_1^2 to intersect IS_1^1 on the Y_f line at G and thus achieve the full-employment output through monetary policy alone.[12] Similarly, through an appropriate expansion of government expenditures, a cut in tax receipts, or some combination of these measures, the IS curve may be shifted from IS_1^1 to IS_1^2 to intersect LM_1^1 on the Y_f line at F and thus achieve the full-employment equilibrium through fiscal policy alone.

Monetary policy may also be used alone to achieve balance-of-payments equilibrium. A sufficiently restrictive monetary policy will shift

and B of Figure 18-5, showed a balance-of-payments surplus or deficit, respectively, instead of the balance-of-payments equilibrium shown by IS_2, LM_2, and BP_2 in Figure 18-8. However, assuming that the IS and LM curves were initially as in Figure 18-5, a decrease in P from P_2 to P_1 would shift the IS and LM curves to IS_1 and LM_1 as in Figure 18-8. The AD curve is determined by the IS and LM curves, not by the BP curve. If there were an initial surplus, a larger decrease in the interest rate than from r_2 to r_1 in Figure 18-8 would then be required to achieve the full-employment equilibrium and balance-of-payments equilibrium. On the other hand, if there were an initial deficit, a smaller decrease, no change or possibly an increase in the interest rate would be required to achieve the two objectives.

[12]Not shown in the figure is the feedback on the IS and BP functions that will result from the higher price level caused by the expansionary monetary policy. In shifting the LM curve to the right, that policy also shifts the AD curve in Part B to the right or up the AS curve to a higher price level. A higher price level will produce a leftward shift of both the IS and BP functions.

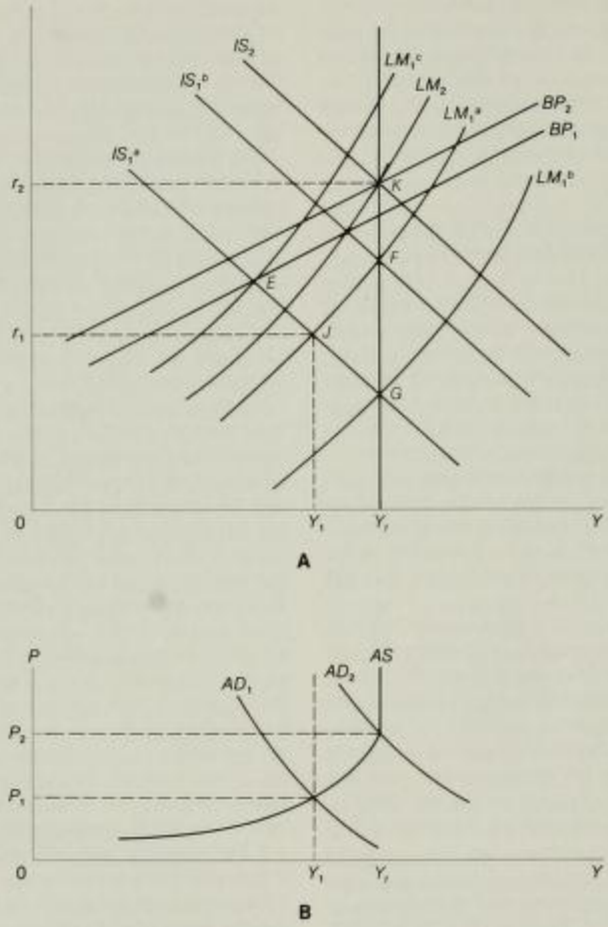


FIGURE 18-9
Balance-of-payments and full-employment equilibria through monetary-fiscal policies

the LM curve leftward from LM_1^a to LM_1^c so that LM_1^c and IS_1^a intersect on the BP_1 curve at E.[13] The deficit in the balance of payments that occurs at the original point J has been removed. However, this has only been done by reducing output below the Y_1 level, which already was below the full-employment level, and is thus not a policy that would likely be adopted. The restrictive monetary policy raises r enough to establish balance-of-payments equilibrium, but the same rise in r reduces interest-sensitive spending enough to cause the indicated reduction in the output level.[14]

One way to achieve both objectives—full-employment equilibrium and balance-of-payments equilibrium—is through coordinated use of both monetary and fiscal policies. Given the point at which the BP function cuts the Y_f line, the dual objective calls for both a higher interest rate and a higher level of income. Expansionary fiscal policy may be relied upon to raise the income level, but the higher income level it produces will in itself reduce the net export balance and make the existing deficit in the balance of payments larger than it was. Some offset to this larger deficit will occur as the expansionary fiscal policy also forces up the interest rate and reduces the net capital outflow. However, a deficit will remain. The answer must be a monetary policy sufficiently restrictive to raise the interest rate to the level at which the net capital outflow and the net export balance that occur at the full-employment income level.

and monetary policy must be sufficiently contractionary to shift LM from LM_1^a to LM_2. These shifts in the IS and LM curves bring about a shift in AD from AD_1 to AD_2 in Part B and a rise in the price level from P_1 to P_2.[15] The IS_2 and LM_2 curves are understood to be those consistent with the new equilibrium price level of P_2. Also the BP curve consistent with the new equilibrium price level is the one labeled BP_2. The higher price level produces a leftward or upward shift of the BP function by reducing exports and increasing imports and tends to increase the size of the deficit. To achieve the dual objectives, fiscal policy must be more expansionary and monetary policy must be more contractionary than would be required in the absence of a rise in the price level. The end product is the balance-of-payments equilibrium and the full-employment equilibrium given by the intersection of IS_2, LM_2, and BP_2 on the Y_f line at K. Output has risen from Y_1 to Y_f, the interest rate from r_1 to r_2, and the price level from P_1 to P_2.[16]

Although Figure 18-9 traces the mechanics of the process by which an appropriate mix of monetary and fiscal policies may achieve the dual objectives of a full-employment internal equilibrium and an external equilibrium, this is all it does. In practice, the process is anything but mechanical. There are limitations on the

[15]Recall that there must be a rise in the price level to

efficacy of monetary and fiscal policies as instruments for achieving full employment in a closed economy and further limitations on the efficacy of monetary and fiscal policies in an open economy as instruments to achieve the dual objectives of full-employment and external equilibrium.

Basically, the policy approach in a balance of payments deficit situation is one that relies on a rise in the interest rate obtained by restrictive monetary policy to meet the external objective and an expansion of aggregate demand by fiscal policy to meet the internal objective. While both objectives may be attained, the very fact that a rise in the interest rate is required to do so introduces other problems. For example, in the judgment of many, the mix of output will be altered in an unfavorable direction. A problem encountered on repeated occasions in the post-World War II period has been the severely depressing effect of a rise in interest rates on construction, especially residential construction. Other investment spending has also been curtailed. Fiscal policy may succeed in expanding output, but because of the rise in interest rates, too little of the additional output may be in capital goods or the kind of goods that increase the economy's long-run capacity to produce. Too much of the additional output may be in public projects with little value beyond the fact that they provide jobs, however important that may be. For monetary and fiscal policies to achieve the most desirable results, economists continue to seek ways to offset consequences such as the one here noted that follows from these policies. For example, one technique tried back in 1961–62 came to be called "operation twist." The Federal Reserve sought to twist the interest rate structure to secure higher short-term interest rates that would induce an inflow of short-term capital from abroad and reduce an existing balance-of-payments deficit, and at the same time it sought to secure lower long-term interest rates that would induce construction and other investment spending to alleviate the recessionary or near-recessionary conditions at home.

Beyond a specific problem like this is a whole range of more general problems faced in the use of monetary–fiscal policies: for example, the difference in the time lags between the date at which monetary and fiscal actions are taken and their impact is felt and variations in the length of these lags from one case to another; the lack of knowledge of what is the proper degree of monetary or fiscal stimulus or restraint in any situation, with the danger of overshooting on the expansion side to bring on inflation or overshooting on the contraction side to bring on recession; the delays faced in getting Congress to act on whatever fiscal actions appear to economists to be best under the circumstances; and so forth. Some problems of this kind are covered briefly in the last two chapters of the book devoted primarily to monetary and fiscal policy.

As long as one recognizes that the application of monetary–fiscal policies faces many difficult problems to which there are anything but perfect solutions, he is not likely to be misled by the simple mechanics of this section and to read into the analysis of the kind presented here more than is in it. However, with these words of caution, the analysis here presented is surely informative and in general indicative of certain basic relationships whose understanding enables one to take his first steps in the macroeconomics of the international economy.

THE EXTENDED INCOME DETERMINATION MODEL: A CONCLUDING NOTE

This chapter concludes the analysis we began far back in Chapter 4. Although any number of related questions have been introduced along the way, e.g., in this chapter the question of simultaneous equilibrium in the balance of payments and the level of output, our basic ques-

tion has been the one posed at the beginning of Chapter 4: What determines the economy's real income or output of final goods and services in any time period? This, as we have seen, is closely related to the question of what determines the level of employment. The further questions of what causes the recurrent short-run fluctuations in the level of output and employment known as business cycles and what explains the long-run rate of growth of output will be discussed in the first three chapters of Part Five.

Clearly, the answer to the basic question of short-run output determination is not a simple thing that we can pack into a summary formulation that is at once neat, brief, and comprehensive. Because our detailed analysis has repeatedly fallen back on simplifying expedients, any summary formulation necessarily compounds the simplification. Recognizing this serious limitation, Figure 18-10 is submitted as a simplified formulation of the more complicated conceptual apparatus developed step by step over these past chapters.

Since our major objective is to explain the determination of aggregate output, the figure is constructed to focus on output—all arrows point in this direction. We see that the immediate determinants of aggregate output and the price level are the aggregate supply and demand functions. In the basic classical theory, the level of output is determined entirely on the supply side. Given that theory's assumptions of a perfectly flexible money wage rate and price level, the aggregate-supply curve is perfectly inelastic at the level of output which can be produced by a fully employed labor force; and the amount of output that can be produced by this number of workers, given their education, training, and skills, depends on the economy's capital stock, state of technology, and supply of natural resources, all of which are fixed in the short run. In the classical case, aggregate demand merely determines the price level at which the full-employment level of output will sell in the market. In Keynesian theory, given its assumption of downward inflexibility of the money wage rate and the price level, the aggregate-supply curve slopes upward to the right over a range of output before becoming perfectly inelastic at the full-employment level of output. With this aggregate-supply curve, aggregate demand affects both the price level and the output level, and if aggregate demand is not adequate, the amount of output produced will be below the amount that provides full employment for the labor force. It is in the Keynesian case that the portion of Figure 18-10 showing unemployment, a difference between the size of the labor force and the number employed, becomes relevant.

In this and other ways, Figure 18-10 has been constructed to fit the Keynesian theory. Thus, it shows the following:

Aggregate-supply function }	Aggregate output, Y Price level, P	{ Aggregate-demand function

in which Y and P are simultaneously determined by the aggregate supply and demand functions. To correspond with the simple classical theory, it would be more accurate to present this part as follows:

Aggregate-supply function }	Aggregate output, Y Price level, P	{ Aggregate-demand function

in which Y, in classical theory the output produced at full employment, is determined on the supply side, and P, the price level at which the full-employment output sells, is determined on the demand side.

In terms of Figure 18-10, the difference between the two theories becomes more striking when we look at the determinants of aggregate demand. In the simple classical theory, most of what is now found to the right of the entry, Aggregate-demand function, disappears; all that remains is the part which shows

Demand for money, M_d
Supply of money, M_s

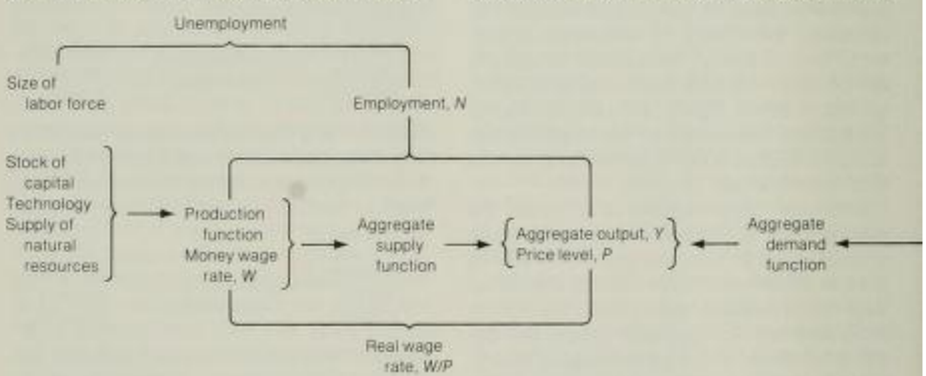
and some of the items under these headings. Beyond this, in the classical theory changes in

aggregate demand result from changes in the supply of money, not from changes in the demand for money. Accepting the demand for money of the classical theory, we have $M_{sp} = 0$ or $M_d = M_t$ and therefore $M_d = k \cdot PY$. To accept that theory is also to accept k as a constant. In equilibrium, $M_s = M_d$ or $M_s = k \cdot PY$. With k constant, PY varies proportionally with M_s. The aggregate-demand side reduces to the one variable, the supply of money, which in conjuction with the constant k establishes the aggregate-demand function.

In contrast, for the extended Keynesian model developed in the last few chapters, all of what is found to the right of the entry, Aggregate-demand function, is needed to explain aggregate demand. We have seen in these chapters that shifts in the aggregate-demand function occur with shifts in either the *IS* function, *LM* function or both, and the approach to the explanation of changes in aggregate demand has been through the various factors that cause shifts in these two functions. Although it could not there be expressed in these terms, the shifts in aggregate demand which were considered in the simple Keynesian model of Part 2 were shifts explained solely in terms of the factors that lie behind the *IS* function. The supply of and demand for money which lie behind the *LM* function were no part of the simple Keynesian model.

Although the interest rate affects I and perhaps also C (as shown by the inclusion of the interest rate among the determinants of both of these spending streams), the interest rate was assumed to be given in the simple Keynesian model. The extended model summarized in Figure 18-10 incorporates the supply of and demand for money, these in turn determine the interest rate, and the interest rate is no longer a given but a variable that affects I and C. The interest rate itself in turn is affected by other determinants of these spending streams and also of the G and $X - M$ spending streams, because changes in any of these determinants affect total spending, which affects the level of income, which affects the demand for money, which affects the interest rate. This is the kind of interaction noted repeatedly in the *IS-LM* analysis of the last few chapters.

Even though we limit ourselves to the determinants specifically listed under the *IS* and *LM* headings, our list is still a fairly long one. The "etc." inserted here and there suggest that it is anything but complete. A rise or fall in the price levels of other countries, technological advances, changes in the distribution of income, money supply, speculative demand for



- IS
  - Consumption, C
    - Disposable income, Y_d
    - Interest rate
    - Price expectations
    - Distribution of income
    - Consumer asset holdings
    - Consumer credit terms
    - Demographic factors
    - Institutional factors
    - etc.
  - Investment, I
    - MEI schedule
      - MEC schedule
        - Expected income from capital goods
          - Businessperson's expectations
          - Technological changes
          - Level of profits
          - Changes in level of output (acceleration principle)
          - Government tax policies
          - Market structures
          - etc.
      - Supply price of capital goods
    - Rate of interest, r
  - Government purchases, G (determined autonomously or in accordance with public policy objectives)
  - Net exports, $(X - M)$
    - Tariffs, quotas, exchange controls
    - Domestic income level
    - Domestic vs. foreign price levels
    - etc.
- LM
  - Demand for money, M_d
    - Transactions demand, Precautionary demand: M_t
      - Level of income, PY
    - Speculative demand: M_{sp}
      - Interest rate, r
  - Supply of money, M_s (determined by monetary authority in accordance with public policy objectives)

FIGURE 18-10
A diagrammatic summary of income theory

money, or consumer asset holdings, expectations of a rising or falling price level, and many other such developments produce shifts in the aggregate-demand function. With a given aggregate-supply function, it is shifts in the aggregate-demand function that bring about changes in output, employment, and the price level. To explain how the overall structure holds

together, let us trace briefly just one of these possible changes through the system.

Suppose the money supply is increased. The rightward shift in the *LM* function will produce some decrease in the interest rate, unless the economy is so deep in depression that it is caught in a liquidity trap. Unless the MEC schedule is perfectly inelastic, there will be some increase in the profit-maximizing capital stock as we move down the MEC schedule as a result of a fall in the interest rate. Unless the MEI schedule is perfectly inelastic (investment already being at that rate at which the supply curve of capital goods is perfectly inelastic), there will be some rise in the rate of investment spending. Unless there is an offsetting decrease in some other component of aggregate spending, there will be a rightward shift in the aggregate-demand function. Unless the economy is at a perfectly elastic or perfectly inelastic portion of the aggregate-supply function, there will be some rise in output, employment, and the price level. If aggregate supply is at the extreme of perfect inelasticity, the increase in the money supply will be purely inflationary; only the price level rises with output and employment unchanged. If aggregate supply is perfectly elastic, only output and employment expand; the increase in the money supply will leave the price level unchanged.

Dozens of other sequences could be traced through this diagram in similar fashion. The diagram, however, serves only as a point of departure. It does little more than lay out in neat but artificial order the major variables with which we must deal. It does not begin to show the complex interrelationships among those variables that we also ignored in the verbal description above. If they were included we would have a bewildering maze of crisscrossing arrows running in all directions. Furthermore, the relationships are dynamic; the time sequence in which successive changes occur should be shown as well. The present diagram stops far short of all this. Nonetheless, it may still be found helpful in tying together the broad dimensions of the theoretical structure developed in the last few chapters.

part five

Cycles, Growth, and Stabilization

chapter 19 Business-Cycle Theory

In the preceding parts of this book, we set as our primary task the development of the theory of the determination of aggregate income and output. This theory basically attempts to do no more than explain how investment, saving, and consumption interact to determine the income level and how changes in these variables produce specific changes in income in a simple, comparative statics framework. A more ambitious theory is one that attempts to analyze the dynamics of income movements, which in a capitalistic economy are seen to pass through alternating periods of expansion and contraction that we call business cycles. A theory capable of doing this goes beyond the theory of income determination—it must somehow answer a whole series of questions that arise specifically from the way the economy behaves during business cycles. Basic among these are questions like the following: Why does the expansion come to an end? Why does the economy move downward into a recession instead of simply leveling off? What determines how far down the economy will go during a slump? Why does the contraction sooner or later end and recovery begin?

Historically viewed, it is difficult to say when systematic study of these and other questions basic to the business cycle began. It certainly did not begin before recognition of the fact that economic activity actually traces out wavelike movements or of the equally important fact that each upswing and downswing in these movements is self-generating. Once a swing begins, it feeds on itself and generates further movement in one direction until it is finally checked and reversed by an accumulation of opposing forces. Clement Juglar, one of the great names in the early study of business cycles, in his *Les Crises Commerciales* (published in 1862) may have been the first to put these features of the economy together.[1] The most commonly held view before this time was that of the economy as a stable mechanism moving along on an even path that was occasionally disturbed by a "crisis." Economists had to progress beyond this kind of view before systematic study of business cycles could be said to have begun.

As study over the years led to an ever better understanding of the nature and causes of business cycles, it also eventually gave rise to the hope that more complete understanding might in time provide the way to altogether eliminate the business cycle or at least sharply reduce its

[1] See J. A. Schumpeter, *History of Economic Analysis*, Oxford Univ. Press, 1954, pp. 1123–24.

severity. In the view of some, a great boost to this hope came with the refinement and application of Keynesian economics in the years after World War II. Keynesian economists came to believe early in the postwar years that fiscal policy could be adjusted as required to prevent sizable cumulative swings in aggregate demand. If this could be done, whatever small variations still occurred in general economic activity might not deserve to be called business cycles. The cycle would have been mastered.

The belief that this could be done was much more common by the late sixties. Many converts were won over by successes like the 1964 tax cut that produced much the results that economists had predicted it would. Increasingly in those years the question was asked: Is the business cycle obsolete? As shown by the chronology of post-World War II business cycles in Part A of Table 19-1 (disregard Part B for the moment), there was an unprecedentedly long period of 105 months of expansion beginning in February 1961. It is understandable that many observors would have been convinced by the late sixties that the business cycle was indeed a thing of the past. During the entire period since 1854 for which business cycles have been dated, the previous record of continuous expansion was the 80-month span from June 1938 to February 1945, and this was hardly a comparable period in view of its domination by World War II. However, as the record shows, the record-long expansion of the sixties did come to an end in November 1969. Part A of Table 19-1 shows that the recession that began then ended a year later in November 1970. This 1969–70 downturn put a stop to the speculation of the sixties that the business cycle was a thing of the past, at least in the minds of most people.[2] Any who still thought otherwise were convincingly disabused of that notion three years later in November 1973 when the U.S. economy, as well as those of western Europe and Japan, entered the most bruising recession since the thirties. For a couple of statistics, from 1973-IV (Roman numerals refer to quarters) to 1975-I, real GNP decreased at a 5.3 percent annual rate; in the worst of these quarters, 1974-IV to 1975-I, real GNP decreased at a 9.9 percent annual rate. Prior to the 1973–75 recession, the one from 1957 to 1958 had the dubious distinction of being the most severe of the post-World War II period and during this one the real GNP declined at an annual rate of 3.3 percent, less than two-thirds of the rate of decline of that in 1973–75.

The month in which the bottom of this most recent recession was reached has been set at March 1975.[3] Accordingly, there have been more than two years of recovery as of the time of this writing in the spring of 1977. During the first year of this recovery, 1975-I to 1976-I, real GNP increased 7.3 percent, a high rate but no higher than the rate to be expected during the first year of recovery from the unusually low

[2]However, an argument could then be made that the 1969–70 recession provided little basis for rejecting the contention that the business cycle was a thing of the past. This recession was unique in that it did not come about despite efforts to prevent it but rather was, in a sense, deliberately planned and executed by the Nixon administration. This is discussed more fully in Chapter 23, but in brief this is what was involved. The administration's attack on the worsening inflation that it inherited was to slow the rate of the economy's advance through restrictive monetary and fiscal policy. This slowing was supposed to somehow reduce the rate of price increase while having little adverse effect on output and employment. However, the nature of the inflation was such that the approach being followed could show some success in checking inflation only by first forcing substantial unemployment on the system. Although this was not publicly declared, it appears that the administration was ready to push these restrictive policies to the point of markedly rising unemployment and recession, if that was the price that had to be paid to check inflation. As it actually turned out, even putting the economy through a recession unfortunately did not have a very noticeable effect on the rate of inflation. This then led the administration in August 1971 to turn to the use of wage and price controls to combat inflation.

[3]The closest thing we have to "official" business cycle dates are those set by the National Bureau of Economic Research, a private organization in New York City which has been engaged in business cycle research for over fifty years. Although the chronology in Part A of Table 19-1 appears in a federal government monthly publication, the source is this private research agency.

level to which the economy had fallen during the recession. However, for the first three quarters of the second year of recovery, 1976-I to 1976-IV, real GNP increased at an annual rate of only 3.6 percent. Measured at annual rates from 1976-I, the growth rates were 4.5, 3.9, and 2.6 percent, respectively, for those three quarters. Declines of growth rates from over 7 percent to those figures would have been less cause for concern if the economy had by then absorbed the slack inherited from the recession, but that clearly had not happened. The unemployment rate during 1976 was 7.7 percent, nowhere near anybody's definition of full employment, and the capacity utilization rate in manufacturing was not much above what it had been during the recession itself. With so great a decline in the growth rate of output in

TABLE 19-1
Turning points and duration of expansions and contractions of classical and growth cycles in the United States, 1945–1975

A.
CLASSICAL CYCLES

DATES OF TURNING POINTS		DURATION IN MONTHS	
PEAK	TROUGH	EXPANSION[a]	CONTRACTION[b]
Feb. 1945	Oct. 1945	80	8
Nov. 1948	Oct. 1949	37	11
July 1953	Aug. 1954	45	13
July 1957	April 1958	35	9
May 1960	Feb. 1961	25	9
Nov. 1969	Nov. 1970	105	12
Nov. 1973	March 1975	36	16

B.
GROWTH CYCLES

DATES OF TURNING POINTS		DURATION IN MONTHS	
PEAK	TROUGH	EXPANSION[a]	CONTRACTION[b]
July 1948	Oct. 1949	—	15
June 1951	June 1952	20	12
March 1953	Aug. 1954	9	17
Feb. 1957	May 1958	30	15
Feb. 1960	Feb. 1961	21	12
April 1962	March 1963	14	11
June 1966	Oct. 1967	39	16
March 1969	Nov. 1970	17	20
March 1973		28	

[a]Measured from trough on preceding line to peak; classical trough preceding Feb. 1945 peak was June 1938; growth cycle trough preceding July 1948 not identified; growth cycle chronology starts with July 1948 and ends with March 1973. As of the time of writing, the National Bureau of Economic Research had not updated its growth cycle chronology, i.e., it had not identified growth cycle turning points since that of March 1973, although at least one more growth cycle trough had clearly been established by early 1975.
[b]Measured from peak to trough on same line.
SOURCE: Classical cycle chronology from *Business Conditions Digest*, U.S. Department of Commerce, November 1976; growth cycle chronology from Prototype Issue, July 31, 1975 draft, of *International Economic Indicators*, National Bureau of Economic Research.

the face of so much slack, fears arose that the economy might be moving toward another recession. However, in the spring of 1977 the outlook improved markedly. The preliminary growth rate of real GNP for the first quarter of 1977 released in June was 6.4 percent at an annual rate, despite the unusually severe winter. The figure would have been well above this under normal winter conditions. The unemployment rate also declined somewhat in the first quarter of 1977 from its 1976 level. In view of these and other developments, the question of whether 1977 would mark the beginning of the U.S. economy's seventh recession since World War II was answered in the negative by most forecasters.

In part the actual answer to this question, which would be provided as the months passed, depends on how one answers an underlying question with which we have not yet dealt: How does one define a recession or contraction, or including expansion as well as contraction, how does one define a business cycle? The traditional definition of a business cycle—the one so far in general use in this country and implicit in the discussion above—runs in these terms: "expansions occurring at about the same time in many economic activities followed by similar general . . . contractions . . ."[4] By this kind of definition, known as the classical definition, a recession or contraction occurs only if there is an absolute fall or a downward change in the direction of overall economic activity that is sustained for some minimum time period. The chronology of business cycles given in Part A of Table 19-1 is one of classical business cycles. Although this definition remains the one usually understood when economists refer to the business cycle in general, another definition has come increasingly into use in recent years. By this second definition, a mere slowing or retardation in the rate at which economic activity is expanding rather than an absolute decline or reversal is sufficient to stamp the period a recession. Because a positive rate of growth is consistent with this definition of recession, a recession of this kind is known as a growth recession, and the business cycle of which it is a part is known as a growth cycle.

The concept of the growth cycle evolved naturally from the kind of economic instability displayed by the U.S. economy and other market-oriented economies during the years since World War II. Unlike the experience of the late nineteenth century through the nineteen-thirties, the more recent kind of instability has been primarily variations in a positive rate of growth rather than reversals in the direction of growth. The underlying upward trend of the economy over the post-World War II years has been sufficiently strong to make absolute downturns much less frequent than during the prewar years. Given this change in the kind of instability exhibited by the economy, the requirements of the classical definition of the cycle makes recessions relatively infrequent. But it does not alter the fact that significant fluctuations in the tempo of economic activity do occur over a period of time that the classical definition simply classifies as a period of uninterrupted expansion. Although these fluctuations are of lesser intensity or severity, they are still fluctuations, and the concept of the growth cycle is needed if these fluctuations are not to be ignored but to be identified as a kind of business cycle.

Part B of Table 19-1 presents a chronology of growth cycles for the United States covering the period for which such a chronology is available. Unlike the chronology of classical cycles, which goes back to 1854, the chronology of growth cycles is limited to the postwar period starting with 1948.[5] For the period covered by both chronologies, it will be seen that the dates of the troughs in some of the classical and growth

[4]A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, National Bureau of Economic Research, 1946, p. 3.

[5]The first growth cycle chronology for the U.S. was provided by I. Mintz in "Dating United States Growth Cycles," *Explorations in Economic Research*, National Bureau of Economic Research, Summer 1974. The "classic" work on the measurement and dating of classical business cycles is A. F. Burns and W. C. Mitchell, *Measuring Business Cycles*, *op. cit.*

cycles are the same. If we select two widely separated troughs common to both chronologies, it will be seen that the number of growth cycles over the period between these troughs is greater than the number of classical cycles, a result which is almost inevitable in view of the difference between the definitions of the two kinds of cycles. A full cycle may be measured from peak to peak or trough to trough, and both chronologies show troughs in October 1949 and November 1970. However, over these twenty-one years, we find four classical cycles and seven or almost twice as many growth cycles. Even more pronounced is the difference between the two chronologies from the February 1961 trough to the November 1970 trough. In the classical chronology, this is one full cycle that contains the record expansion of 105 months, from February 1961 to November 1969, that gave rise to so much talk in the late sixties about the obsolescense of the business cycle. But look at the same period, February 1961 to November 1970, in the growth cycle chronology. Instead of one full cycle, there are *three* full cycles in the identical time span. There was sufficient retardation in the growth of economic activity from April 1962 to March 1963 and from June 1966 to October 1967 to inject two contractions into a period that was one of uninterrupted expansion according to the classical definition.

One cannot help but speculate on the difference in outlook that would have existed during the sixties among economists and business people if they had then been accustomed to using the term business cycle to mean growth cycles rather than classical cycles and if they had then had available the chronology of growth cycles that did not appear until the seventies. Under these circumstances, would a book entitled *Is the Business Cycle Obsolete?*[6] have appeared in 1969? Of more recent interest, would economists and business people at the beginning of 1977 have been talking of the danger of a recession ahead or would they have been talking of an existing recession? The retardation in the growth of economic activity may have been sufficient to identify a growth recession starting early in 1976 and running to the beginning of 1977. Whether this is indeed the case must await the evaluation of the evidence by the economists at the National Bureau of Economic Research who establish the turning point dates for the growth cycles as well as the classical cycles.

The likelihood is that over the years economists and others will come to think more and more in terms of growth recessions and growth cycles, despite a long tradition in which the term, business cycle, was automatically understood to mean the classical cycle. Not that we have heard the last of the classical cycle: The 1973–75 recession eminently qualified as a classical recession and the 1970–75 cycle of which it was a part similarly qualified as a classical cycle. Both classical and growth cycles will almost surely continue to occur. Arthur F. Burns, a leading scholar in the study of business cycles early in his career, wrote some lines over thirty years ago that may be almost as true now as they were then.

> For well over a century business cycles have run an unceasing round. They have persisted through vast economic and social changes; they have withstood countless experiments in industry, agriculture, banking, industrial relations, and public policy; they have confounded forecasters without number, belied repeated prophecies of a "new era of prosperity" and outlived repeated forebodings of "chronic depression." Men who wish to serve democracy faithfully must recognize that the roots of business cycles go deep in our economic organization, that the ability of government to control depressions adequately is not yet assured, that our power of forecasting is limited, and that true foresight requires policies for coping with numerous contingencies.[7]

[6] M. Bronfenbrenner, editor, *Is the Business Cycle Obsolete?* Wiley-Interscience, 1969.

[7] A. F. Burns, "Stepping Stones Toward the Future," *27th Annual Report of the National Bureau of Economic Research*, 1947, p. 27, reprinted in A. F. Burns, *The Frontiers of Economic Knowledge*, National Bureau of Economic Research, 1954, p. 45.

The business cycle has persisted over the years since World War II much as Dr. Burns suggested would be the case at the start of the postwar period. However, from the perspective of the classical cycle, the postwar record as a whole, including the grinding 1973–75 experience, has been one of shorter and less severe downturns than occurred over the century to which Dr. Burns refers. The explanation for this is found in part in a factor alluded to earlier: The development of economic knowledge since World War II has increased the ability of economists to prescribe policy that will prevent the downswings that occur from becoming more serious than they otherwise would, but it has not done more than this. Another factor has been the various institutional and structural changes in the economy that make for a more stable system, perhaps the most important of which is the introduction of automatic stabilizers, something we will look at in Chapter 24.

Whatever progress has already been made or will be made in the effort to moderate the business cycle rests basically on the ability of economists to understand how the economy generates these cyclical contractions and expansions. This involves questions like those asked in the first paragraph of this chapter, questions to whose answers economists have made important contributions over the years. These contributions are the concern of the present chapter, but the list of these economists is so long and some of the analyses so involved that any attempt to provide even the briefest discussion would require a full volume.[8] If we narrow our sights to the last few decades, most economists would include in their lists of important contributions the work of Schumpeter, Hansen, Metzler, Harrod, Kalecki, Samuelson, Kaldor, Hicks, Goodwin, and Duesenberry.[9] In this chapter we will do no more than touch on the contributions made by a few of these modern economists. In the three sections that follow, we will be concerned specifically with Kaldor's model, Samuelson's multiplier–accelerator interaction, and Hicks's model of the cycle.

The variability of investment spending plays an important role in all modern business cycle analyses, and some of our work in Chapters 11 and 12 on investment will be carried over to this chapter and incorporated with modifications into the business cycle analyses presented here. In those earlier chapters our purpose was simply to consider what determines the level of investment spending, approaching investment spending as one component of aggregate spending. Thus, in the first part of Chapter 11 we developed the profits theory of investment. Because profits depend on the level of income and because, according to the profits theory, investment depends on the level of profits, we could express that theory in terms of the investment equation $I = I_a + eY$. The general idea that investment depends on the level of income will, with substantial modification, be found to be a cornerstone of Kaldor's theory of the cycle that we will

[8] The "classic" among the books of this kind is G. Haberler, *Prosperity and Depression*, Harvard Univ. Press, 1958. See also A. H. Hansen, *Business Cycles and National Income*, Norton, 1964 and R. A. Gordon, *Business Fluctuations*, second edition, Harper, 1961. A current textbook in the area is C. A. Dauten and L. M. Valentine, *Business Cycles and Forecasting*, 4th edition, South-Western, 1974.

[9] Only a single work of each of these economists will here be noted: J. A. Schumpeter, *Business Cycles*, McGraw-Hill, 1939; A. H. Hansen, *Business Cycles and National Income*, Norton, 1964; L. A. Metzler, "Nature and Stability of Inventory Cycles," in *Review of Economics and Statistics*, Aug. 1941, pp. 113–29, reprinted in R. A. Gordon and L. R. Klein, eds., *Readings in Business Cycles*, Irwin, 1965, pp. 100–29; R. F. Harrod, *The Trade Cycle*, Clarendon, 1936; M. Kalecki, *Essays in the Theory of Economic Fluctuations*, Allen & Unwin, 1939; P. A. Samuelson, "Interaction Between the Multiplier Analysis and the Principle of Acceleration," in *Review of Economic Statistics*, May 1939, pp. 75–78, reprinted in G. Haberler, ed., *Readings in Business Cycle Theory*, Irwin, 1944, pp. 261–69; N. Kaldor, "A Model of the Trade Cycle," in *Economic Journal*, March 1940, pp. 78–92, reprinted in his *Essays on Economic Stability and Growth*, Free Press, 1960, pp. 177–92; J. R. Hicks, *A Contribution to the Theory of the Trade Cycle*, Oxford Univ. Press, 1950; R. M. Goodwin, " "The Nonlinear Accelerator and Persistence of Business Cycles," in *Econometrica*, Jan. 1951, pp. 1–17; and J. S. Duesenberry, *Business Cycles and Economic Growth*, McGraw-Hill, 1958. For further references, see the extensive bibliography in J. J. Clark and M. Cohen, eds., *Business Fluctuations, Growth, and Economic Stabilization*, Random House, 1963, pp. 623–69.

consider in the first of the following three sections of this chapter.

In the last part of Chapter 11 we turned to the acceleration principle as one of the leading relationships advanced by economists to explain the level of investment spending. There we had the investment equation $I_t = I_a + w(Y_t - Y_{t-1})$. Starting with a nonmathematical formulation of an idea provided by Alvin H. Hansen, Samuelson worked out the mathematics to show how the interaction between the acceleration principle and the multiplier analysis could trace out cyclical movements in the level of income. Samuelson's mathematical statement is not really a business-cycle model, for it does not go much beyond the mere mechanics of the interaction. His analysis, however, has become a classic in business-cycle literature, for it showed how the interaction, once set into motion, is capable, in and of itself, of generating an endless series of successive expansions and contractions in the income level. Since continuous fluctuations in income are in fact characteristic of a free economy, it is not surprising that students of the business cycle soon thereafter seized on the interaction of the multiplier and accelerator as a factor that helps to explain the business cycle. The mechanics of the process by which the interaction generates fluctuations will be traced in the second of the following sections of this chapter.

A number of business-cycle models came to be built on the multiplier–accelerator interaction, but the one that attracted the most attention is that developed by Hicks. The final section of this chapter will be devoted to a sketch of the major features of the Hicksian model of the business cycle.

KALDOR'S MODEL OF THE CYCLE

This model of the cycle may be approached as a logical extension of the model of income determination that was developed in connection with the profits theory of investment in Chapter 11.[10] There we introduced the investment function $I = I_a + eY$. The saving function, as in the first model presented in Chapter 4, was $S = S_a + sY$. Graphically the model appeared as in Part A of Figure 19-1. The equilibrium level of Y is Y_e, the only income level at which planned saving and planned investment are equal. With any given pair of linear saving and investment functions, there is a single equilibrium position, and any disturbance that results in a shift in either function or in both would tend to be followed by a movement to a new equilibrium position. However, from the particular viewpoint of the business cycle, this model offers little; it indicates more stability than the real world seems to show.

In Chapter 11 and now here, we implicitly assume that MPS > MPI, or $s > e$. Graphically, the assumption is that the slope of the saving function is greater than that of the investment function and that the investment function therefore cuts the saving function from above. Actually, the MPI may very well exceed the MPS. In this case, illustrated in Part B of Figure 19-1, there is again a single equilibrium position, but it is now an unstable one. Any disturbance producing a movement above Y_e means that $I > S$ and that the income level would rise without limit, first to full employment and then beyond to hyperinflation; any disturbance producing a movement below Y_e means that $S > I$ and that the income level would collapse to zero output and employment. In contrast to the case shown in Part A of the figure, the case shown in Part B gives us a greater instability than the real world shows. As an approach to an explanation of the business cycle, both cases are found wanting, one for too much stability and the other for too little.

[10] See pp. 186–86.

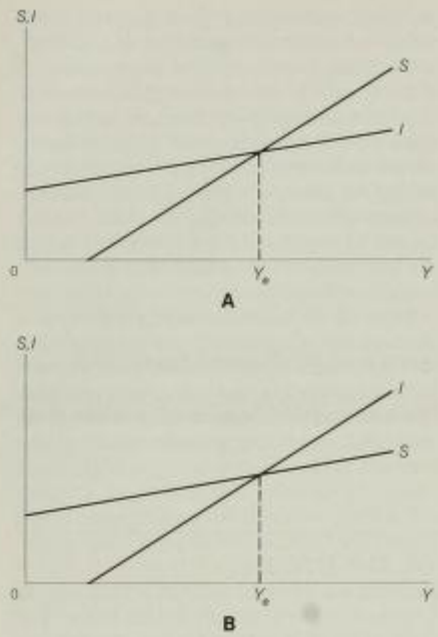


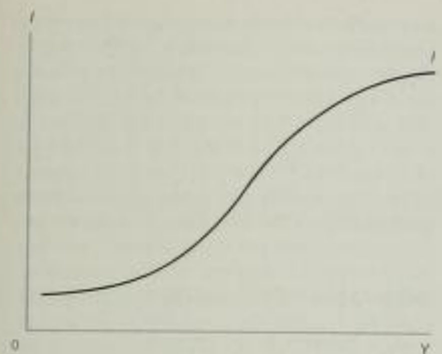
FIGURE 19-1
Stable and unstable equilibria

Kaldor concludes from the preceding analysis that the saving and investment functions cannot both be linear, at least not over the full range of income change that occurs during the business cycle. Furthermore, some good reasons can be advanced to support the argument that neither of the two functions is indeed linear.

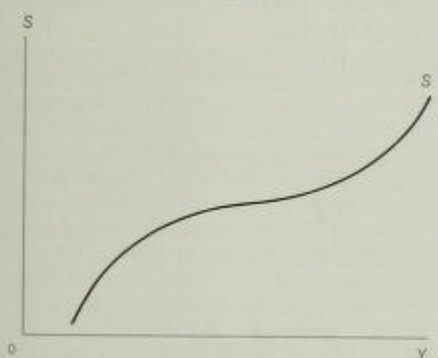
If we think of the cycle as divided into phases of relatively high, relatively low, and "normal" income, we would not expect the MPI to be the same under all three conditions, which is the assumption of the linear investment function. A nonlinear function that appears to conform more closely with the behavior of investment during the course of the cycle is that shown in Part A of Figure 19-2. Here, for both relatively high and relatively low levels of income, the curve is approximately flat—the MPI is approximately zero. The MPI may be expected to approach zero at relatively low income levels, for such income levels are accompanied by conditions of large excess capacity. Due to the presence of excess capacity, the rise in income from its low point will not at first induce investment spending. In other words, there is a range of income over which successive increments to income, ΔY, will be accompanied by small or zero increments to investment, ΔI, or $\Delta I/\Delta Y$ will be very small or zero over this range of Y. In the opposite case of relatively high levels of income, Kaldor argues that the MPI will be similarly small "because rising costs of construction, increasing costs and increasing difficulty of borrowing will dissuade entrepreneurs from expanding still faster—at a time when they already have large commitments."[11]

Similarly, in the case of the saving function we would not expect the MPS to be the same under the varying income conditions specified above. A nonlinear saving function that appears to conform more closely with the behavior of saving during the course of the cycle is shown in Part B of Figure 19-2. At relatively high and relatively low income levels, the situation is the reverse of that shown for investment. Here the MPS is relatively large compared with its magnitude at "normal" income levels. This may be explained along the following lines. As income falls to relatively low levels in recession, people cut saving drastically in order to maintain something like their previous standard of living; then, in the early stages of recovery, they increase saving sharply in order to restore the depleted savings to their usual level. The MPS is thus relatively large, and the saving function is correspondingly steep. In the opposite case, at which income rises to relatively high levels, people tend to save not only a larger amount

[11] *Op. cit.*, p. 180.



A



B

FIGURE 19-2
Nonlinear investment and saving functions

but also a larger proportion of their income. This tendency is reinforced by a shift in the distribution of income toward profits and away from wages, given the fact that the MPS of profit recipients is higher than that of wage recipients. These forces are reflected in a steepening of the saving function at relatively high income levels in comparison with its slope over the range of "normal" income.

Figure 19-3 is derived by merely combining the nonlinear investment and saving functions of Figure 19-2.[12] This figure shows multiple equilibria, with both A and B as stable positions. At income levels below Y_1 or between Y_2 and Y_3, $I > S$, so the income level rises; at income levels between Y_1 and Y_2 or above Y_3, $S > I$, so the income level falls. C is an unstable position, so the income level Y_2 is not a stable equilibrium level. If Y is between Y_2 and Y_3, it will rise to Y_3; if it is between Y_1 and Y_2, it will fall to Y_1. It appears that the economy can reach stability only at some particular high level of income, Y_3, or at some particular low level of income, Y_1.

What we now have, however, is anything but a model of the business cycle, for the cycle is made up of alternating expansions and contractions, and Figure 19-3 simply shows two positions of stable equilibrium. For the answer to this, let us quote Kaldor: "The key to the explanation of the trade cycle is to be found in the fact that each of these two positions is stable only *in the short period:* that as activity continues at either one of these levels, forces gradually accumulate which sooner or later will render that particular position unstable."[13] If it can indeed be shown that the stable equilibrium at A becomes unstable over time and forces a movement to B, and similarly in the opposite case, we will have pushed ahead from what is shown by Figure 19-3 to a model of the business cycle.

Figure 19-3 shows how saving and investment may vary with the level of income over the course of the cycle. The level of income, however, is not the only major force affecting the amount of the economy's saving and investment. Kaldor introduces a second variable that plays a major role in cyclical changes in saving and investment—this is the economy's capital

[12]We follow Kaldor's model by assuming that the two functions are nonlinear as described by Figure 19-2. However, the analysis that follows would remain valid even if only one of the two functions was as described above and the other was linear.

[13]*Op. cit.*, p. 182.

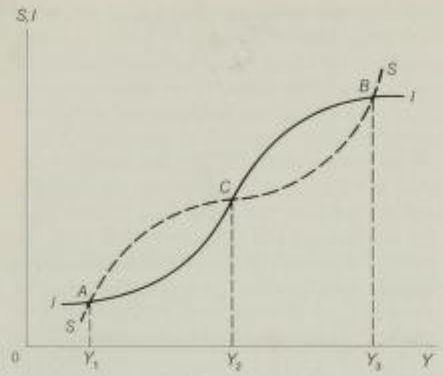


FIGURE 19-3
Multiple equilibria

stock. Saving is a direct function of the capital stock: For any level of income, the greater the capital stock, the larger is the amount of saving. Investment is an inverse function of the capital stock: For any level of income, the greater the capital stock, the smaller is the amount of investment. Our next step is to trace out how the changes in the capital stock that occur over time alter the equilibrium situations derived in Figure 19-3.

In Figure 19-4, the Stage I diagram corresponds to Figure 19-3. We start off in this diagram with the assumption that the economy is at the equilbrium position shown as *B*. This corresponds to a relatively high, or above "normal," income at which investment is correspondingly high. But the higher the rate of investment, the more rapid is the increase in the size of the capital stock. In terms of the analysis of Chapter 10, a growing capital stock means a movement down the MEC curve, which, unless somehow offset, in turn means a downward shift in the MEI curve, which appears here as a downward shift in the *I* curve. At the same time, the growth in the capital stock, which is a growth in the economy's wealth, will tend to push up the saving curve, *S*. This is a rise in the average propensity to save induced by an increase in the economy's wealth. As shown by the Stage II diagram, the downward movement of the *I* curve and the upward movement of the *S* curve result in a gradual shift to the left in the position of *B* and a gradual shift to the right in the position of *C*, so *B* and *C* are brought closer to each other. The critical point is reached when these gradual shifts of the *I* and *S* curves make the two curves tangential and bring *B* and *C* together as in the Stage III diagram. Now, at the position of $B + C$, $S > I$ in both directions, and the equilibrium is unstable in a downward direction. The cyclical contraction, once started, reduces the income level until a new stable equilibrium is reached at the relatively low level that corresponds with *A*.

Again, *A* is a stable equilibrium only in the short run. Over time the *S* and *I* curves gradually shift; but now, with the system at a relatively low income level, the *I* curve shifts upward and the *S* curve shifts downward, as shown by the Stage IV diagram. If the level of investment corresponding to *A* is less than replacement requirements, some upward shift in the *I* curve will occur sooner or later for replacement reasons alone. Apart from or in addition to this, over time investment opportunities gradually accumulate. In terms of the analysis of Chapter 10, the MEC curve shifts to the right, pushing up the MEI curve, which here would mean an upward shift in the *I* curve. At the same time, any decline in the capital stock or in the economy's wealth that occurs during the period of relatively low income will tend to lower the average propensity to save, or push the *S* curve downward. These shifts cause the position of *A* to move to the right and that of *C* to move to the left, thus bringing *A* and *C* closer together, as is evident from the Stage IV and Stage V diagrams. Again, the critical point is reached when these gradual shifts of the *I* and *S* curves make the two curves tangential and bring *A* and *C* together, as in the Stage VI diagram. The $A + C$ position is unstable in an upward direction, since $I > S$ on both sides of the position. The cyclical expansion, once started, raises the in-

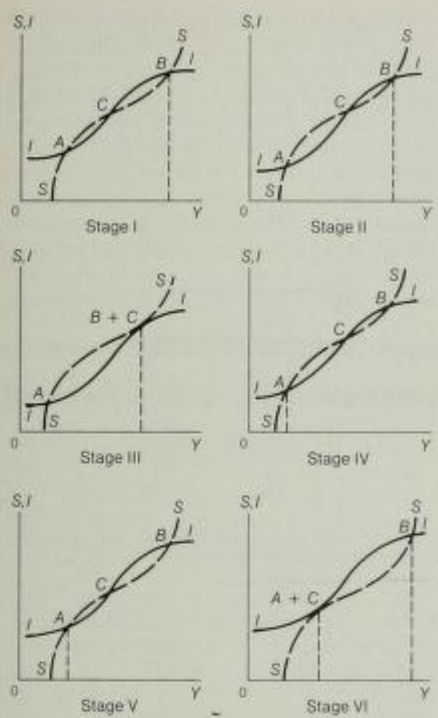


FIGURE 19-4
The Kaldor model

come level until a new state of equilibrium is reached at the relatively high level that corresponds with *B*. The curves then gradually return to the positions shown in the Stage I diagram, and another cycle begins.

The cyclical process described by Kaldor is thus self-generating. The very movement to relatively high income levels brings into play forces that, after a period of time, produce a downward movement to relatively low income levels, and vice versa. These forces, such as the changing size of the average propensity to save and the accumulation and decumulation of capital that occur over the cycle, are inherent in the economic process; they are endogenous forces in the full sense of that term.

Kaldor's theory of the cycle appeared just four years after the *General Theory* and is a relatively simple and very neat theory built directly on Keynes's saving-investment analysis. Although Keynes devoted a chapter of the *General Theory* to "Notes on the Trade Cycle" and offered further discussion of the subject at other points, he did not develop a theory of the cycle as such. His book was concerned basically with the development of what he regarded as a new theory of the determination of the *level* of income, and the theory of *fluctuations* of income received only passing attention. The model of income determination developed in connection with the profits theory of investment in Chapter 11 of this book is taken from Keynes. It is interesting to note that Kaldor's theory of the cycle emerges essentially from the substitution of his particular nonlinear saving and investment functions for the linear ones in this specific income model and from his ingenious tracing of the implications that follow from the quite different saving and investment relationships given by the nonlinear functions. If Keynes had gotten to the point of developing a formal model of the cycle in the *General Theory*, it might well have been much like the one examined in this section.

SAMUELSON'S MULTIPLIER–ACCELERATOR INTERACTION

In Chapter 11 we described the acceleration principle as a relationship in which any time period's net investment, I_t, is made dependent on the value of the accelerator, w, and the

change in income between that and the preceding period, $Y_t - Y_{t-1}$.[14] Our interest there was limited to an explanation of investment spending, and the acceleration principle was considered solely from this point of view. In the present section we retain the acceleration principle as an explanation of investment spending and combine it with the familiar multiplier analysis. The interaction between the two produces the following general sequence of events.

Assuming initially a change in Y, the investment of any period, I_t, will differ from that of the preceding period, I_{t-1}, as indicated by the acceleration principle. However, a rise or fall in the level of investment will affect the level of Y in succeeding periods in a way determined by the multiplier analysis. But any change in the level of Y from one period to the next must, again via the acceleration principle, affect investment in each following period. Thus, as was shown by Samuelson in the article noted earlier, there is the possibility of an endless sequence of changes in income and output and in consumption and investment as a result of the way in which the multiplier and the acceleration principle interact. In what follows we will trace this interaction in detail through a multiplier–accelerator model.

A Multiplier–Accelerator Model

The accelerator investment function we used in Chapter 11 was $I_t = I_a + w(Y_t - Y_{t-1})$. Although not noted there, this function is but one of a group of possible investment functions incorporating the acceleration principle. Instead of making the accelerator apply to the change in total output, $Y_t - Y_{t-1}$, another function in this group makes the accelerator apply only to the change in the output of consumer goods, $C_t - C_{t-1}$. The resulting function, $I_t = I_a + w(C_t - C_{t-1})$, is used by Samuelson and will be used here.

[14]See pp. 188–94.

Still other functions in this group result from employing a lagged form of the acceleration principle instead of the unlagged form used above. Thus, if we assume that the nonautonomous portion of investment spending in any period depends, not on the change in aggregate output between the current and the preceding period, but on the change in aggregate output between the preceding period and the period before that, we have as our investment function $I_t = I_a + w(Y_{t-1} - Y_{t-2})$. Or, if we use lagged consumption instead of lagged income, we have $I_t = I_a + w(C_{t-1} - C_{t-2})$. The lagged function appears to be the more plausible. A rise in the current period's spending over the preceding period's spending may indeed call forth business decisions to expand investment, but the actual investment spending cannot occur simultaneously with the appearance of the increase in spending that calls it forth. The lagged function allows for an interval of time between the decision to invest and the actual investment expenditures.

Similarly, in the case of the consumption function, we may work with lagged or unlagged relationships. The consumption function employed in earlier chapters, $C_t = C_a + cY_t$, assumes that consumption of each period depends on that period's autonomous consumption and that period's income level. It assumes that consumers adjust their spending to changes in income as fast as those changes occur. One alternative to this function is the lagged function $C_t = C_a + cY_{t-1}$, which assumes that consumption in each period depends on that period's autonomous consumption and the preceding period's income level. It allows an interval of time between the receipt of income and the expenditure of that portion the recipient chooses to spend.

There are clearly many other possible investment and consumption functions, based on different assumptions concerning the timing and the way in which consumption and investment spending are influenced by changes in income and other variables. In what follows, since our

immediate purpose is simply to illustrate the interaction between the multiplier and the acceleration principle, we will work with only one of the possible pairs of functions and will use the one that corresponds with that used by Samuelson in his original article:

$$C_t = C_a + cY_{t-1}$$
$$I_t = I_a + w(C_t - C_{t-1})$$

Since income and output of any period equals consumption plus investment spending during that period, or since $Y_t = C_t + I_t$, by substitution we derive the following general equation:[15]

$$\boxed{Y_t = C_a + cY_{t-1} + I_a + w(C_t - C_{t-1})}$$

This equation states that aggregate income and output in any period equals the sum of autonomous consumption, C_a, and autonomous investment, I_a, plus an additional amount of consumption that depends on the marginal propensity to consume, c, times the income of the preceding period, and an additional amount of investment that depends on the capital–output ratio or the accelerator, w, times the change in consumption between the current and the preceding period. In other words, if we know C_a and I_a for any period and C for the current and preceding periods, we can, given the values of c and w, determine the income and output of any period by substituting in the equation for Y given above.

Table 19-2 shows the results that follow over 22 successive time periods from substituting in this equation. The column headings of the table are set up to match the order of the components given in the equation for Y. Total consumption for any period is the sum of columns (2) and (3), total investment for any period is the sum of columns (4) and (5), and the income for any period as given in column (6) is therefore the sum of columns (2) through (5). In this illustration, the MPC is assumed to be 0.6 [column (3)] and the accelerator to be 1.5 [column (5)]. Column (8) shows the change in income from the table's first period, period 1, to every other period. Note that in the first two periods, 1 and 2, the economy is in equilibrium with the level of income and output unchanged at 100.

If we now break into the table in period 3, induced consumption, cY_{t-1}, is seen to be the same as in period 2, since Y in period 2 is the same as Y in period 1 (both are 100). Induced investment, $w(C_t - C_{t-1})$, in period 3 is zero, since C in both periods 3 and 2 is 70 and $70 - 70$ is zero.[16] Now suppose that I_a rises in period 3 from 30 to 40 and C_a remains unchanged at 10. How will the rise in I_a in period 3 affect the level of income in that period? Since induced consumption in period 3 depends on income of period 2, which is the same as income of period 1, induced consumption remains at 60 in period 3 as it was in period 2. Since induced investment in period 3 depends on the change in C between periods 3 and 2, which is zero, induced investment also remains unchanged at zero. The rise in income and output in period 3 therefore equals the rise of 10 in autonomous investment spending. We are assuming that this rise in autonomous investment spending is permanent, so I_a remains at 40 in period 4 and subsequent periods. In-

[15] In what follows we will work with the equation in this form. However, since $C_t = C_a + cY_{t-1}$ and $C_{t-1} = C_a + cY_{t-2}$, we may substitute these expressions for C_t and C_{t-1} in the equation above and write that equation in the alternative form

$$Y_t = C_a + cY_{t-1} + I_a + wcY_{t-1} - wcY_{t-2}$$

or

$$Y_t = C_a + I_a + c(1 + w)Y_{t-1} - wcY_{t-2}$$

Expressed in this way, investment is shown to depend on Y lagged plus the values of c and w instead of on C unlagged plus the value of w.

[16] In developing the profits theory of investment in Chapter 11, we used the term *induced investment* to describe investment that is dependent on the absolute level of a variable, using income for that variable. In the acceleration principle, investment is dependent on the change in the level of a variable, which in the present illustration is consumption. Because investment that depends on a change in the size of a variable is quite different in meaning from investment that depends on the absolute size of a variable, separate terms would seem to be preferable. However, the general practice is to use "induced investment" to cover both cases, so the meaning in any case must be inferred from the context in which it is used.

TABLE 19-2
A multiplier–accelerator model

(1) PERIOD	(2) AUTONOMOUS CONSUMPTION C_a	(3) INDUCED CONSUMPTION $c(Y_{t-1})$ $c = 0.6$	(4) AUTONOMOUS INVESTMENT I_a	(5) INDUCED INVESTMENT $w(C_t - C_{t-1})$ $w = 1.5$	(6) INCOME AND OUTPUT Y	(7) CHANGE IN INCOME AND OUTPUT FROM PRECEDING PERIOD	(8) CHANGE IN INCOME AND OUTPUT FROM PERIOD 1
1	10	60.0	30	0	100.0		
2	10	60.0	30	0	100.0	0	0
3	10	60.0	40	0	110.0	10.0	10.0
4	10	66.0	40	9.0	125.0	15.0	25.0
5	10	75.0	40	13.5	138.5	13.5	38.5
6	10	83.1	40	12.1	145.2 (P)	6.7	45.2
7	10	87.1	40	6.1	143.2	−2.0	43.2
8	10	85.9	40	−1.8	134.1	−9.1	34.1
9	10	80.5	40	−8.2	122.3	−11.8	22.3
10	10	73.4	40	−10.7	112.7	−9.6	12.7
11	10	67.6	40	−8.6	109.0 (T)	−3.7	9.0
12	10	65.4	40	−3.3	112.1	3.1	12.1
13	10	67.2	40	2.8	120.0	7.9	20.0
14	10	72.0	40	7.2	129.2	9.2	29.2
15	10	77.5	40	8.2	135.7	6.5	35.7
16	10	81.4	40	5.9	137.3 (P)	1.6	37.3
17	10	82.4	40	1.5	133.9	−3.4	33.9
18	10	80.3	40	−3.1	127.2	−6.7	27.2
19	10	76.3	40	−6.0	120.3	−6.9	20.3
20	10	72.2	40	−6.2	116.0	−4.3	16.0
21	10	69.6	40	−3.9	115.7 (T)	−0.3	15.7
22	10	69.5	40	−0.3	119.2	3.5	19.2

duced consumption now rises from 60 in period 3 to 66 in period 4, since income of period 3 has risen to 110 from its level of 100 in period 2. Induced investment, 0 in period 3, rises to 9 in period 4, for the change in C from period 3 to 4 is 6. Thus, substituting in the equation for Y gives us the following:

$$Y_t = C_a + cY_{t-1} + I_a + w(C_t - C_{t-1})$$

For Period 3:

$$110 = 10 + 60 + 40 + 1.5(70 - 70)$$

For Period 4:

$$125 = 10 + 66 + 40 + 1.5(76 - 70)$$

For Period 5:

$$138.5 = 10 + 75 + 40 + 1.5(85 - 76)$$

All subsequent changes in induced consumption, induced investment, and the level of income may be traced through period by period in this same fashion.

The data given in Table 19-2 show that the level of income rises to a peak (P) in period 6, then falls to a trough (T) in period 11, rises to another peak in period 16, and falls to another trough in period 21. In short, the level of income and output traces out a series of cyclical fluctuations. It is interesting to note that these fluctuations are generated in this model by nothing more than one change—the rise in autonomous investment spending to a higher but constant rate in periods 3, 4, and subsequent periods. Thus, by incorporating the accelerator in the investment function, a permanent increase in autonomous investment no longer necessarily means that the level of income and output will simply move up to a higher level and reestablish equilibrium there; instead it may, as we have seen, set the system off on a series of self-generating cyclical fluctuations without establishing a new equilibrium.

Variations on the Model

If the figures of Table 19-2 were continued to period 23 and beyond, one would discover additional cycles, but cycles whose amplitude would become smaller over time. In the absence of any further disturbance, such as a change in C_a or I_a, these cycles would eventually die out, and the level of income would stabilize at 125, the equilibrium level indicated by the multiplier alone. This result, a series of "damped" cycles, follows from the specific values assumed for c and w in Table 19-2. Other values for c and w will give different results.

In his article, Samuelson set forth the various patterns of change that the level of income might display for different combinations of values of c and w, given a change in autonomous spending that upsets a previous equilibrium. All possible patterns fit into one of four cases, or five if we include a special case: (A) income moves upward or downward at a decreasing rate and asymptotically approaches a new equilibrium; (B) income fluctuates through a series of cycles of smaller and smaller amplitude until the cycles virtually disappear (as in Table 19-2); (C) income fluctuates through a series of cycles of wider and wider amplitude; (D) income moves upward or downward at an increasing rate; (E) income fluctuates through a series of cycles of constant amplitude (the special case). Cases A and B resemble each other in their stability—both tend to converge toward an equilibrium. Cases C and D resemble each other in their instability—both tend to diverge from equilibrium by increasing amounts. Case E is the in-between situation—movement is neither toward nor away from equilibrium. Otherwise classified, Cases B and C exhibit cyclical fluctuation; Cases A and D do not. The special case, E, goes with B and C in this classification. Each of these five income patterns is illustrated in Parts A through E of Figure 19-5.

We have noted that the differences between these five income patterns result solely from different combinations of values for c and w. Every possible combination of c from zero to 1.2 and of w from zero to 5 may be located in Figure 19-6.[17] Those combinations of c and w that fall in the region labeled B will produce

[17] This figure is adapted from Samuelson's on p. 268 of Haberler, *Readings in Business Cycle Theory*.

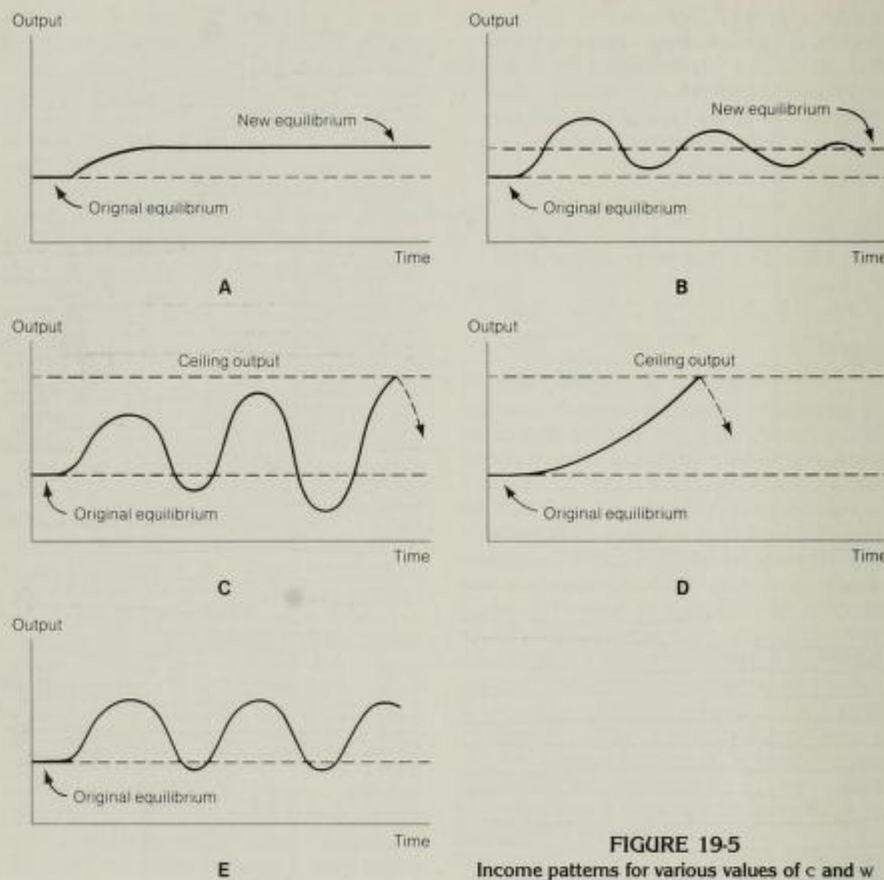


FIGURE 19-5
Income patterns for various values of c and w

a series of damped cycles such as those shown in Part B of Figure 19-5 and in Table 19-2. Similarly, those combinations of *c* and *w* that fall in regions *A*, *C*, and *D* will produce income movements such as those of Parts A, C, and D of Figure 19-5. Combinations of *c* and *w* that fall along the broken line marked *E* produce cycles of constant amplitude such as those of Part E of Figure 19-5.

Figure 19-6 also indicates that any particular income pattern, for example the explosive case of region *D*, can result either from a relatively low multiplier and a relatively high accelerator (e.g., 2.5 and 5) or from a higher multiplier and a lower accelerator (e.g., 10 and 2.5).[18] Simi-

[18] It may be noted that the same explosive case will follow even with an accelerator of zero if the MPC is greater than 1. This is similar to one of the theoretical possibilities examined in the development of the multiplier. See p. 76.

larly, the damped-cycle case that results from a combination of c and w falling in region B is possible with a relatively low multiplier and a relatively low accelerator (e.g., the values in Table 19-2, 2.5 and 1.5) or with a higher multiplier and a still lower accelerator (e.g., 5 and 0.5).[19]

In the development of the multiplier in Chapter 5, we saw that the size of the multiplier determines how much income will rise or fall in response to a given increase or decrease in autonomous investment or consumption spending. Beyond this, all that could be said at that point was that a high multiplier would, given the volatility of autonomous investment spending, produce a more unstable system than a low multiplier. However, with the addition of the acceleration principle, we can say something more definite about the path of change that will be followed by income over time, ranging from the explosive path of region D at one extreme to the equilibrium path of region A at the other. Also, despite the fact that the volatility of aggregate investment spending springs in part from the volatility of autonomous investment spending, which is not explained by the acceleration principle, we can say that some part of the overall volatility of investment spending is explained by the internal workings of the multiplier–accelerator interaction. As was illustrated in Table 19-2, all that is needed to set the interaction into motion is a change in autonomous investment (or autonomous consumption) spending. Given such a change, the system itself generates a greater or lesser volatility of induced investment as part of its own self-contained process of change.

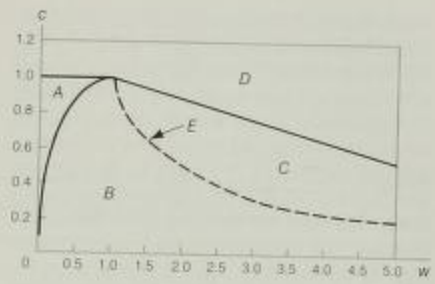


FIGURE 19-6
Boundaries of regions for values of c and w yielding different income patterns

The Multiplier–Accelerator Model and the Business Cycle

All students of the business cycle agree that movements in economic activity that are to be called business cycles take the form of recurrent alternations of expansion and contraction. A cumulative movement in one direction over time creates pressures that eventually cause a reversal and a movement in the opposite direction. The duration and amplitude of periods of expansion and contraction may vary substantially from cycle to cycle, and the duration of successive full cycles also shows wide variation. However, there remains as the essence of the cycle the recurrent alternations of upward and downward movements in economic activity that may be called oscillations.

From this point of view, the movements of economic activity (measured by aggregate output) shown in the various patterns of Figure 19-5 do not all display the oscillatory movements characteristic of the business cycle. In

[19]It should be noted that the boundaries of regions shown in Figure 19-6 pertain specifically to the equation of the type above: $Y_t = C_a + cY_{t-1} + I_a + w(C_t - C_{t-1})$. Any other equation will give us a different set of boundaries. Thus, if we merely substitute $(C_{t-1} - C_{t-2})$ in the preceding equation—or, in other words, apply w to lagged rather than unlagged consumption—the instability of the model will be reduced. A figure like Figure 19-6 prepared for this equation would, among other differences, have the line E lying higher, indicating that higher values for c and w would be needed to produce antidamped, or explosive, cycles. For another case, if, instead of making the accelerator apply only to changes in consumption, we made it apply to changes in total output, we would make for a more unstable system. Thus, if we should substitute $(Y_{t-1} - Y_{t-2})$ in the equation above, the line E in Figure 19-6 would shift downward, actually to a vertical position at $w = 1$, so any value of the accelerator greater than 1 would lead to an antidamped, or explosive, cycle.

this connection, the noncyclical nature of Case A stands out most clearly. Given a disturbance, the system moves smoothly from an original equilibrium to a new equilibrium with no oscillation whatsoever. There is simply nothing cyclical here. Case B produces oscillations, but they are damped oscillations that tend to die out over time and thus conflict with the historical record that shows no tendency for cycles to disappear. However, the tendency for the cycle to disappear as portrayed by Case B follows from the assumption of only a single disturbance—for example, a one-time spurt in autonomous investment. In this event, the amplitude of the oscillations shrinks over time to zero. But in reality, further disturbances can be expected to occur quite frequently and at random intervals.[29] Technological advances, innovations, wars, natural disasters, can all produce jolts to the system. The inherent tendency of the system to react cyclically to a disturbance will result in an oscillatory movement, even though the disturbances occur at random over time. The resultant cyclical pattern will be quite irregular in terms of duration and amplitude of cycles, but the pattern in the real world is, of course, far from regular. Thus, we may say that what otherwise shows up as a tendency for the cycle to disappear in Case B may be converted into an unending sequence of cycles by the addition of randomly distributed erratic shocks to the system.

Case C produces oscillations that increase in amplitude, or become explosive. This pattern at least produces the continuing oscillatory movement characteristic of business cycles, but it conflicts with the observed fact that actual movements are not explosive. However, a model with values for the multiplier and accelerator that puts the system into the region of Case C may be made consistent with the nonexplosive nature of observed business cycles if so-called *buffers* are added to the system. We will discuss these in more detail in the next section on Hicks's model; here it is sufficient to note that such buffers impose an upper limit, or "ceiling," on the expansion and a lower limit, or "floor," on the contraction and thus convert what otherwise would be the explosive upward or downward movements called for by the model into the limited oscillations characteristic of the actual business cycle.

What was said here about buffers in Case C applies also to Case D, except for the fact that in the latter case there is a directly explosive movement that is restrained instead of oscillations of increasing amplitude that must be eventually restrained. This may be seen in Cases C and D of Figure 19-5, where a ceiling is shown. Values for the multiplier and accelerator that put the system into the region of Case D can thus produce cycles, again if buffers are added to the model as in Case C and if reasons are given for the system to reverse direction after striking the ceiling. Some reasons why reversal may be expected are found in Hicks's theory of the cycle that we turn to in the next section.

In the special situation of Case E, the cycle neither dies out nor explodes but goes on indefinitely at a constant amplitude. The problem here is the implausibility of this case—the result follows only in the very special situation in which the product of the accelerator and the marginal propensity to consume equals 1, a combination unlikely to be realized in practice. Any departure from this combination puts the system into the region of Case B or C, with the results that follow as discussed above.

The various types of movement that result from the multiplier–accelerator interaction have provided the basis for two groups of business-cycle theories. In one group are the theories of economists like Hicks and Goodwin, who have built theories on the assumption that the values of the multiplier and accelerator are such as to produce antidamped, or explosive, cycles. Hansen, the major name on the other side, has

[29] This idea of erratic shocks was first brought forward by R. Frisch, "Propagation Problems and Impulse Problems in Dynamic Economics," in *Economic Essays in Honour of Gustav Cassel*, Allen & Unwin, 1933, pp. 171–205.

been the proponent of the *weak accelerator* theory. In his view, the interaction produces only damped cycles. The preponderance of opinion appears to support the antidamped cycle, and in the next section we turn to a discussion of the best known of the cycle models that proceed along these lines—the model developed by Hicks.

HICKS'S THEORY OF THE BUSINESS CYCLE

Business cycles, when viewed as fluctuations in the economy's real output of goods and services, have historically appeared as movements above and below a rising line of trend or growth. In this sense, it is a commonplace that the real output of the economy in a recession may be substantially larger than that same economy's output at the peak of the prosperity of a cycle a decade or so earlier. The mere growth of the economy makes this so.

Hicks begins by recognizing that since cycles have historically appeared against this background of growth, a theory of the cycle should similarly be constructed against a background that provides a theory of growth. Hicks defines a long-run equilibrium growth path for the economy that is determined by the growth rate of autonomous investment. The ratio of equilibrium income to autonomous investment depends on the magnitude of the accelerator and the multiplier, a result drawn from the Harrod-Domar theory of growth, which we will examine in Chapter 21. Hicks also assumes that autonomous investment tends to grow at a fairly constant percentage rate over the long run. With a fairly stable accelerator and multiplier, it then follows that there is an equilibrium growth path for income that exhibits the same constant percentage rate of growth as autonomous investment. The failure of actual output to move along this equilibrium growth path over time, its tendency instead to run now above it and then below it, identifies cycles. Hicks's theory then focuses on these movements that constitute cycles, but they are in this way kept in the perspective of a theory of the underlying growth path of the economy.

Hicks uses the term *autonomous* investment in somewhat the same sense that we did in earlier chapters to mean "public investment, investment that occurs in direct response to inventions, and much of the 'long-range' investment . . . which is only expected to pay for itself over a long period."[21] The remainder of all investment, and an amount that makes up a large proportion of the total of net investment in ordinary circumstances, is so-called *induced* investment, or investment that is called forth, directly or indirectly, by past changes in the level of output. Hicks recognizes that the distinction between autonomous and induced investment is not sharp in practice, but it is a distinction that is critical in developing his theory of the cycle.

We can begin to work our way into that theory through Figure 19-7. The line *AA* shows autonomous investment and the line *EE* the equilibrium growth path for income that is based on line *AA*. The vertical axis is in logarithmic scale, so that the straight lines given in the figure indicate a constant percentage rate of growth for the variables in question. Over time, equilibrium income given by line *EE* remains a constant multiple of autonomous investment, the size of the multiple depending on the magnitude of the accelerator and the multiplier, as noted above. The assumption that autonomous investment grows at anything like a constant rate

[21]*Op. cit.*, p. 59.

is, of course, unrealistic, and Hicks's model permits occasional shifts in the rate of growth of autonomous investment. However, from the point of view of his model, variations in autonomous investment are exogenous changes (changes arising as a result of forces outside the model), whereas the basic purpose of the model itself is to investigate how the system it describes will behave in the absence of such exogenous changes.

The line *CC* shows the path of growth of ceiling output that can be produced with full employment of the economy's resources. It is not absolutely necessary that this line lie above *EE*, but this is one possibility and the one considered here. The remaining line, *FF*, describes the path of the lower limit, or floor, to which real income can fall during the contraction phase of the cycle. There will be more to say about this ceiling and floor in what follows, but first we must note why the income level does not follow the equilibrium growth path described by *EE*. Then, given a departure from this path, we will see why the lines *CC* and *FF* set the upper and lower boundaries within which the cycle moves.

The ordinary picture of the business cycle is, of course, one in which a period of expansion is followed by a period of contraction and this in turn by a period of expansion, and so forth. Although one also speaks of a long-run equilibrium growth path, the idea that the economy moves along in anything like a state of equilibrium is a denial of the existence of the cycle. However, to break into the apparatus described by Figure 19-7, we assume that over a period of time ending at point *a* the economy has been in equilibrium or that the level of output has been expanding along the *EE* path and is at point *a*. We assume next that an innovation or some other development produces a temporary spurt in autonomous investment, after which autonomous investment returns to its previous path described by line *AA*. This is the same technique we described in the Samuelson interaction to get the system out of equilibrium,

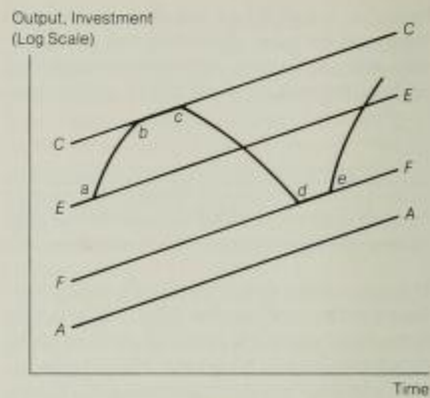


FIGURE 19-7
The Hicksian cycle

although we assumed there that the new higher level of autonomous investment was sustained. Once given this spurt in autonomous investment, the path followed by the income level depends on the magnitude of the accelerator and the multiplier. Hicks believes that in the standard business cycle the magnitudes of these parameters are such as to fall in regions *C* or *D* of Figure 19-6, so any divergence of income from the *EE* path will take the form of oscillations that either increase in amplitude (*C*) or simply produce an explosive upward movement without oscillation (*D*). His study of the economic history of the past century and a half leads him to the conclusion that the movement of the economy over this long period cannot be explained as a damped-cycle movement (region *B* in Figure 19-6) that was kept going by a series of random shocks. In his view, this long period is organized into a series of seven- to ten-year cycles that display a regularity that cannot be accounted for by such an erratic-shock, damped-cycle mechanism. Having thus taken the position that the underlying structure is one that produces antidamped, or explosive, cycles, Hicks then brings the ceiling to output

(*CC*) into the model to explain why explosions do not take place. Combining the floor (*FF*) with the ceiling sets the limits within which cyclical movements occur.

Thus, starting at point *a* and assuming a temporary burst of autonomous investment spending, there is set into motion a cumulative movement upward that is explained by the multiplier–accelerator interaction. The actual movement of the economy's output is indicated by the path of the line from point *a* to point *b*. The economy's ceiling output is also growing with time, but with the rate of growth of actual output exceeding the rate of growth of ceiling output, the rate at which actual output is growing must sooner or later be checked by bumping against the ceiling. This brings into play a crucial element in the analysis. In Hicks's words, "When the path has encountered the ceiling, it must (after a little) bounce off from it, and begin to move in a downward direction. This downward movement is inevitable."[22] In other words, the expansion must end and a contraction must follow.

Simply stated, the inevitability of the contraction follows from the impossibility that output can continue to rise at a rate faster than that at which the output ceiling itself is rising. Depending on the particular induced investment equation one works with, induced investment in any one period is made to depend in some specified way on the change in output over preceding periods. However, the maximum change in output between periods once the ceiling has been hit must be less than the change between periods during the time interval that the economy is moving upward toward the ceiling. In other words, once the system is at point *b*, the growth of output per time period must be less than it was per time period from point *a* to point *b*. In a related context, an examination of the multiplier–accelerator interaction in Table 19-2 shows that the mere decrease in the amount of increase in output between periods is sufficient to bring about a decrease in total output, or a downturn from the peak level of output attained. There is a lag between the time the ceiling is struck and the time the downturn begins, a lag assumed in Figure 19-7 to equal the time interval involved in moving from point *b* to point *c*, but sooner or later the downturn in output must begin. The inevitability of this downturn requires nothing more than the working of the acceleration principle combined with the assumption of a sufficiently strong accelerator to cause the system to hit the ceiling, once it departs from the equilibrium path of *EE*.

Once the downturn begins, the accelerator starts to work in the opposite direction. The change in income is now negative, so induced investment begins to decline. On the upswing, the limit to the expansion of real investment was set by the capacity of the system to produce, but on the downswing the limit to negative investment (disinvestment) in any period is set by that period's depreciation. For example, the decline in output may be such that the acceleration principle calls for disinvestment of $30 billion in a particular period, but if the wear and tear on the capital stock is only $10 billion for the period, then $10 billion is the maximum disinvestment for the period. Businessmen collectively have no way of disinvesting at the desired higher rate. Under conditions such as these, the acceleration principle becomes inoperative in the downward direction. What happens to the income level during the slump is then determined by the simple multiplier theory without any accelerator at work. The line *FF* is, in other words, set by autonomous investment times the multiplier.

Hicks holds that autonomous investment will typically decline during the course of a slump (the *AA* line shifts downward) but will remain positive and, over time, still maintain a more or less steady rate of growth. Thus, as shown by the movement from point *d* to point *e*, income may move upward along the floor as the growth of autonomous investment continues to support income but at a slump level. How long

[22] *Ibid.*, p. 98.

income may move along the floor depends in part on the redundancy of capital inherited from the splurge of induced investment in the earlier periods. Once this is worked off, new investment orders begin to appear for replacement purposes, and gross investment moves above the level accounted for by autonomous investment. We have a spurt in investment spending that pushes the income level upward from the *FF* line. Once this income movement begins, the accelerator again comes into play, induced investment appears, and a cumulative upward process is under way. The rise in income does not stop with the *EE* line, the income level consistent with the *AA* line for autonomous investment. Via the interaction of the multiplier and accelerator, income overshoots the *EE* line and goes on up until finally restrained by the ceiling, the *CC* line, from which, as before, it bounces off and starts the downward movement of another cycle.

The next cycle would follow the course of its predecessor, in line with the essentials of Hicks's theory covered in the preceding pages. We will not further pursue this theory, nor will we go far into the criticisms of it, although there have been many. Different critics focus on what they regard as the major deficiencies of Hicks's theory, but they all pay some attention to the fact that the model is founded on the acceleration principle in a rigid form. For example, to Kaldor this is the major weakness of the model.[23] If the acceleration principle in rigid form is found to be unacceptable, then the multiplier–accelerator interaction that incorporates it and is at the heart of the Hicksian model is also unacceptable. As Duesenberry has put it, "The basic concept of multiplier–accelerator interaction is an important one but we cannot really expect to explain observed cycles by a mechanical application of that concept,"[24] and this is what Hicks's model appears to try to do.

Kaldor, whose cycle theory was discussed earlier, makes no use of the acceleration principle. In his model, investment is related directly to the *level* of income and inversely to the stock of capital. This approach, which is also associated with names like Kalecki and Goodwin, breaks the unrealistic, inflexible tie of investment to changes in output that is implied by the rigid acceleration principle, but it still retains the basic idea of the acceleration principle. Instead of the investment function incorporating the strict acceleration principle,

$$I_t = I_a + w(Y_{t-1} - Y_{t-2})$$

this approach gives us

$$I_t = I_a + gY_{t-1} - jK_t$$

where K is the stock of capital at the beginning of period t and where g and j are constants. The new equation simply says that if output, Y, increases while the capital stock, K, remains constant, investment will rise to enlarge the capital stock, other things being equal. If, on the other hand, the capital stock increases while output remains constant, investment will fall as the desired stock of capital is reached. Thus, the new equation still includes the basic idea of the acceleration principle, for investment in any period will still aim to bring the actual capital stock into line with the amount desired for the production of the level of output that was recently turned out.[25] The difference is that this approach does not contain any direct reference to the rate of change of income and output over successive periods. Although it retains a link between changes in output and the level of

[23] N. Kaldor, "Hicks on the Trade Cycle," in *Economic Journal*, Dec. 1951, pp. 833–47, reprinted in *Essays on Economic Stability and Growth*, Free Press, 1960, pp. 193–209.

[24] J. S. Duesenberry, "Hicks on the Trade Cycle," in *Quarterly Journal of Economics*, Aug. 1950, pp. 464–76, reprinted in J. J. Clark and M. Cohen, *op. cit.*, pp. 322–42.

[25] The new equation becomes identical with an earlier one in the special case in which g equals the capital-output ratio, w, and j equals unity. Adopting these values for the coefficients (and for simplicity setting I_a equal to zero),

$$I_t = wY_{t-1} - K_t$$

In the same way,

$$I_{t-1} = wY_{t-2} - K_{t-1}$$
$$wY_{t-2} = K_{t-1} + I_{t-1}$$

Since K_{t-1} equals the capital stock at the beginning of period $t - 1$ and since I_{t-1} is the change in capital stock

investment, it does this in such a way as to avoid some of the shortcomings of the rigid acceleration principle.[26]

---

during period $t - 1$, the capital stock at the beginning of period t is the sum of $K_{t-1} + I_{t-1}$, or K_t, or

$$wY_{t-2} = K_t$$

Substituting wY_{t-2} for K_t in the first equation above, we have

$$I_t = wY_{t-1} - wY_{t-2}$$

or

$$I_t = w(Y_{t-1} - Y_{t-2})$$

which is one of the expressions for the acceleration principle that was earlier noted.

[26]See R. C. O. Matthews, *Business Cycles*, Univ. of Chicago Press, 1959, Chs. 2 and 3, and J. S. Duesenberry, *Business Cycles and Economic Growth*, Ch. 3.

Hicks's use of the acceleration principle in its rigid form is but one of a number of criticisms that have been made of his model, but we will not go into the others here. However, it may be noted that every economist who has criticized one aspect or another has also highly praised this work of Hicks. For example, Kaldor refers to the "many brilliant and original pieces of analysis" found in Hicks's theory, and Duesenberry describes it as an "ingenious piece of work" and "the first coherent theory of the cycle to appear in some years." Despite the fact that the model has now been around for over twenty-five years and despite the amount of other work that has been done in business-cycle theory during this period, Hicks's model remains, in a sense, the last word in business-cycle theory.

chapter 20

The Nature of Economic Growth

The fluctuations in economic activity that we call business cycles historically take place around a rising line of trend or growth. It is an economic fact of life that cycles and growth go together in this way, and there is a mutual conditioning of each by the other over time. Nonetheless, we still find that few models are concerned with both cycles and growth. Thus, the simple multiplier–accelerator model of the preceding chapter is capable of producing either growth or oscillations but not both together. For those values of *c* and *w* that generate oscillations, we get only oscillations that are cyclical in origin. A model of this kind that included a theory of growth would show that these oscillations are also affected by growth and that growth in turn is affected by the cycles.

Since the builders of models are all aware of the interdependence between growth and cycles in the actual world, their separation in economic theory is, as might be guessed, due to the difficulty of devising a theoretical apparatus capable of handling both together. A model that does do this is appropriately described as a model of *cyclical growth*, and, as Professor Allen has said, "It does seem that, to get cyclical growth, we need a model of quite considerable complexity."[1] Such models are beyond the level of this book. Accordingly, we will limit ourselves to several models of economic growth that are essentially limited to a theory of growth. This parallels the treatment in the preceding chapter, in which the models were essentially limited to a theory of the business cycle.[2]

Growth models as such attempt to explain what determines the rate at which the economy's real output grows over a period of time that is long relative to the period of the business

[1] A model of cyclical growth seeks to explain growth as well as the cycles that occur during growth and is not to be confused with the growth cycle noted in the preceding chapter. The theory of growth cycles does not concern itself with why the economy shows growth over the long run but takes this underlying growth as given and seeks to explain only those recurring retardations in growth which constitute growth cycles. The quotation in the text is from R. G. D. Allen, *Macro-Economic Theory*, St. Martin's Press, 1967, p. 385. A model of cyclical growth manageable to the nonmathematical reader is that by R. M. Goodwin, "A Model of Cyclical Growth," in E. Lundberg, ed., *The Business Cycle in the Post-War World*, Macmillan, 1955, reprinted in R. A. Gordon and L. R. Klein, eds., *Readings in Business Cycles*, Irwin, 1965, pp. 6–22.

[2] Although Hicks's theory of the cycle is built on the foundation of a growth theory, there is not the formal integration of cycle and growth theory that is needed to constitute a theory of cyclical growth.

cycle, and beyond this they consider what is required in order that this growth rate be that rate at which the economy progresses along a path of steady, full employment. We will defer the analysis of growth models to the following chapter in order to devote this chapter to some background on the meaning and measurement of growth, the growth record of the U.S. economy, the relationship between growth and environmental quality, and estimates of the various sources that account for the growth record of this country.

THE MEANING AND MEASUREMENT OF ECONOMIC GROWTH

Economic growth can be most simply defined as the increase in the economy's output. From the classical definition of the business cycle, since business-cycle expansions are, in general, periods during which output increases, and business-cycle contractions are periods during which output decreases, growth by this definition appears to be something we can observe from one quarter to the next during expansions, and the absence of growth appears to be what we observe during contractions. Viewed in this way, economic growth and business-cycle expansions become almost the same thing. This is essentially what the press means by growth when it reports the latest quarterly increase in real GNP as a measure of the economy's "growth" during that three-month period.

Most economists use a definition of growth like the above, but the time period regarded as relevant to the concept of growth is at the very minimum the length of a full business cycle. It then follows that if the level of the economy's output were to remain unchanged from one business-cycle peak to the next, the economist would find zero growth for that cycle. The expansion phase of the cycle, during which output increases, would not register growth but would, in this case, be the phase during which the economy did no more than regain the ground lost during the preceding contraction. In this view, the economy would show growth only to the extent that the later peak lies above the preceding peak.[3] This is far more reasonable than calling any observed increase in output growth even when that increase still leaves aggregate output below the level reached at an earlier date.

We will use here the increase in the economy's output as the basic definition of growth, working with the version that makes the business cycle the minimum time period over which to measure growth. Our best measure of the economy's output is real GNP, or GNP in constant dollars. The reason for specifying constant dollars is, of course, that changes over the years in GNP in current dollars are the result of a mixture of price changes and output changes. Therefore, if growth is defined as the expansion of the economy's output and if we are to use GNP as a measure of growth, then price changes must be removed from GNP as in the constant-dollar series. Furthermore, if we are interested not merely in how much the economy's aggregate output expands over time but in how much the amount of output produced per person expands over time, real GNP must also be corrected for population increases. These two adjustments will give us what is for many purposes the most useful concept of economic growth—expansion in real GNP per capita.

From the basic concept of real GNP per capita, various related concepts of economic growth have been derived, each of which em-

[3]For the actual record, Table 20-2, which appears on p. 389, shows growth rates for each full cycle since 1910. In each cycle, it may be noted, output at the later peak was greater than at the earlier peak, as is indicated by the positive growth rate shown for each cycle.

phasizes a different facet of growth. One is "real consumption per capita," a concept that indicates the growth in consumer economic welfare. Such figures are derived initially from the deflated data for the personal consumption expenditures component of GNP converted to a per capita basis and sometimes adjusted for the changes in leisure that may be regarded as a part of real consumption. A further adjustment could be made by adding to personal consumption expenditures that portion of deflated government purchases of goods and services that is "clearly" in the nature of public consumption, such as public parks and recreation. The problem encountered in making this adjustment, is, of course, the fact that a large part of government purchases cannot be classified as "clearly" of a consumption or of an investment nature. A second and quite different growth concept is "real GNP per unit of labor input" or per combined unit of labor and capital input, a concept that stresses the changes in the economy's efficiency or productivity over time. These are but two of a number of growth concepts, each of which emphasizes a particular feature of the economy over time.

Aggregate real GNP or per capita real GNP is initially recorded as a time series of dollar amounts. In order to simplify comparisons between different time periods and between different economies, growth is usually expressed as an average annual rate between the beginning and terminal years of each period selected rather than in terms of the absolute dollar change over the period. Thus, for the U.S. economy, real GNP per capita (in 1972 prices) increased from $2,584 in 1929 to $5,869 in 1973, an increase of 127.1 percent in a 44-year period. Expressed as an average annual rate of growth, real GNP per capita increased at a 1.8-percent rate over this period.[4] Notice, however, that when measured from the depression level of $1,769 in 1933, the rise to $5,869 over the years to 1973 represents an increase of 217.7 percent, or an average annual rate of growth of 2.9 percent, far greater than the growth rate found when 1929 was taken as the beginning year.

This illustration demonstrates how the choice of beginning and terminal years can markedly affect the rates derived. When a severely depressed year such as 1933 is used as the beginning year, the rate of growth to 1973 is very impressive. When the prosperous year of 1929 is taken as the beginning year, which thereby also includes the depression years between 1929 and 1933, the rate is not too much more than half of the other. To obtain the fairest representation of the actual growth rate over a period, the beginning and terminal years of the period should be similar in terms of the business cycle phase.[5] For this purpose, the unemployment percentage is probably the best guide. Ap-

[4]The average annual growth rate may be determined from a compound interest table by finding the interest rate required to produce the percentage change in the indicated time period. For the present example, the rate required for $1 to grow to $2.271 in forty-four years will be a rate between two of the rates given in the headings of such a table. In the ordinary compound interest table, interpolation is then required to find the growth rate to the nearest tenth of a percent. This need for interpolation may be avoided by using a specially prepared growth-rate conversion table from which the growth rate to the nearest tenth of a percent may be read off directly. (See *Long Term Economic Growth*, Bureau of Economic Analysis, U.S. Dept. of Commerce, 1973, pp. 114–43.)

Alternatively, one can directly compute the growth rate by using the compound interest formula:

$$GNP_t = GNP_b(1 + r)^n$$

where t designates the terminal year, b the beginning year of the period, n the number of years in the period, and r the average annual rate to be determined. Solve by converting the equation to logarithms:

$$\log(1 + r) = \frac{(\log GNP_t - \log GNP_b)}{n}$$

Substituting the data given in the text, r is found to be 1.8 percent. For 1976, real GNP per capita was $5,879 which gives an r of 1.7 percent for the period 1929–76. If our example had been in terms of real consumption per capita, the initial equation would have been

$$C_t = C_b(1 + r)^n$$

which may be solved in the same way.

[5]A way around dissimilar business cycle phases is to adjust the actual beginning and terminal figures to what they would have been if the economy had been operating at potential output each year. See footnote 18 on pp. 394–95. An obvious problem with this, however, is that of accurately

proached in this manner, the growth rate for the 1933–73 period is seen to be misleadingly high, because in 1933 the unemployment rate was at the all-time peak of 24.9 percent. In contrast, the unemployment rate in 1929 was only 3.2 percent, well below but at least within range of the 4.9-percent rate of 1973. Thus, for the 40-to-45 year period ending in 1973, the growth rate based on 1929 is clearly more realistic than that based on 1933 or any of the deeply depressed years between 1929 and 1933.

By the same criterion, for the slightly longer period ending in 1975, neither 1929 nor 1933 nor any other of the years during the thirties would be an appropriate beginning year. The unemployment percentage in 1975 was 8.5 percent, very high relative to 1929 but still low relative to 1933 or any other year during the thirties. If one wants a growth rate for a 40-to-45 year period ending with 1975, he has no choice but to use one of the years from 1930 to 1935, but the rate so computed will be misleadingly high. Using the unemployment percentage as a criterion, 1941 with its 9.9 percent unemployment figure is the most distant year (disregarding pre-1929 experience) that is at all comparable in this regard.

The misleading growth rates that can result from an economically inappropriate choice of beginning and terminal years and from the business-cycle fluctuations that may dominate any short period of years tend to be evened out when the growth rates are computed for a longer period. In the following section we will examine growth rates for the U.S. economy over both long periods and shorter periods. The shorter periods have been selected to begin and end with the same phase of the business cycle.

GROWTH RECORD OF THE U.S. ECONOMY[6]

Table 20-1 summarizes the growth record of the U.S. economy for the period 1839–1959 and for three 40-year subperiods into which this long period has been divided. The record is truly impressive for the overall period of 120 years. There is probably no other economy in history that has done as well during any period of equal length. Over this 120-year period, the aggregate output of the U.S. economy grew at a 3.66-percent average annual rate, which means that it doubled, on the average, every 19 years, or that there was a 32-fold expansion of output in five such intervals, or in less than a century. Allowing for population growth of almost 2 percent per year, per capita GNP grew at 1.64 percent per year over this long period, or per capita income doubled every 43 years.

When the long period is broken down into 40-year subperiods, we find that the annual growth rate of aggregate GNP declined in each successive subperiod—from 4.31 to 3.72 to 2.97 percent. Notice that the same decline does not appear in the per capita data because of variations in the rate of population increase. Of the three subperiods, aggregate GNP showed the most rapid growth during the 1839–79 period. Per capita GNP showed the slowest rate of growth during this period, however, because of the relatively rapid rate of population growth

---

estimating what that potential output would have been in each year. This problem is briefly considered in Chapter 24 in connection with the concept of the full employment budget surplus.

[6]The Department of Commerce periodically undertakes a comprehensive revision of the national income and product accounts. The last two such revisions were those published in the August 1965 and the January 1976 issues of the *Survey of Current Business*. None of the references cited in the following pages reflect the most recent revision. However, the differences between the prior and revised estimates of constant-dollar GNP do not alter substantially the long-term growth trends. For a couple of examples, the average annual growth rate for 1929–64 was raised from 1.6 to 1.7 percent by the 1965 revision, and the rate for 1948–74 was left unchanged at 3.6 percent by the 1976 revision. See *Survey of Current Business*, August 1965, p. 20, and January 1976, Part I, p. 25.

TABLE 20-1
Growth rates and price changes, 1839–1959

	ENTIRE PERIOD	FORTY-YEAR SUBPERIODS		
	1839–1959	1839–1879	1879–1919	1919–1959
Aggregate real GNP	3.66%	4.31%	3.72%	2.97%
Price level	1.15	−0.16	1.91	1.40
Population	1.97	2.71	1.91	1.30
Real GNP per capita	1.64	1.55	1.76	1.64

SOURCE: R. W. Goldsmith, "Historical and Comparative Rates of Production, Productivity, and Prices," Joint Economic Committee, *Employment, Growth, and Price Levels, Hearings*, Part 2, 1959, p. 271. (All rates are average annual rates.)

during this period—the rate was more than twice that of the most recent 40-year period. Though the rate of growth of aggregate output was less during the 1879–1919 period than during the preceding period, population growth slowed even more, to account for the highest of the three subperiod figures for per capita GNP. A still slower rate of population growth in the most recent subperiod resulted in a small decrease in the per capita GNP growth rate, even though the growth rate of aggregate GNP fell off sharply between these two subperiods.

Just as the rate of growth for this long period conceals variations in its subperiods, so each subperiod conceals even greater variations in the still shorter subperiods into which it may be divided. Table 20-2 divides the period 1910–73 into such shorter subperiods, each of which covers the span of a full business cycle measured from peak to peak.[7] The growth rate of aggregate real GNP (not per capita GNP) and the implicit price deflators are given for each of these business cycles.[8]

Over the fourteen business cycles of this sixty-four year period, the highest growth rate, 7.1 percent occurs during the cycle that includes the years of World War II, and the highest peacetime growth rate, 5.4 percent, occurs during the cycle of the early twenties. The slowest growth rate, 0.1 percent, occurs during the cycles that include the Great Depression and the readjustment following World War II. The relatively high rate of 4.4 percent during the 1948–53 cycle reflects the impact of the Korean War and the still unsatisfied demand pent up from World War II. Following the relatively low rates of 2.8 and 2.7 during the two cycles from 1953 to 1960, the rate during the long cycle from 1960 to 1969 rose to 4.3 percent, and then fell off to 3.4 percent in the most recent cycle from 1969 to 1973.

The experience during the fifties and sixties illustrates well the huge difference in real output that results from a slow growth rate like 2.7 or 2.8 percent and a more rapid rate like 4.3 per-

[7] In contrast to the limitation of Table 20-2 to years of business-cycle peaks, detailed tables called growth-rate triangles, each including about 3,000 entries, have been prepared that show the growth rate of aggregate real GNP for every possible combination of initial and terminal years for the years from 1890 to 1969. (See *Long Term Economic Growth*, pp. 105–10.)

[8] The growth rates and the deflators given are actually those based on real GNP of the private domestic economy. This excludes GNP originating abroad and GNP originating with government (government employee wages, salaries, and supplements) and runs about 90 percent of aggregate real GNP over time. The difference between growth rates computed on the two bases becomes appreciable only during wartime and postwar adjustments because of the large rise and fall in federal government expenditures for national security during these periods. For example, including government GNP, the growth rate for the 1937–44 cycle is 8.5 instead of 7.1 percent and for the 1944–48 cycle is −2.4 instead of 0.1 percent.

cent. If the actual growth rate during the 1960–69 cycle had remained at the level of the preceding two cycles, the *cumulative* loss in output (valued in 1972 dollars) over the nine years from 1961 through 1969 would have approached three quarters of a *trillion* dollars. The loss in 1969 alone would have been about \$140 billion—instead of the actual GNP of \$1,078.8 billion (in 1972 dollars) recorded for that year, the figure would have been about \$940 billion. The difference between 2.7 or 2.8 percent and 4.3 percent is, of course, quite large—the latter is, respectively, 60 and 54 percent greater than the former two figures—but the results that follow from smaller differences are still dramatic. If the growth rate during 1960–69 had jumped only to 3.5 percent, a 25-percent increase over a 2.8-percent rate, it would still have meant almost an extra \$300 billion (in 1972 prices) of goods and services for the nine-year period. Figures like these make apparent why economists can get concerned about changes as small as 1/10 of a percentage point in growth rates.

TABLE 20-2
Growth rates and price changes during U.S. business cycles, 1910–1973

BUSINESS CYCLE (PEAK TO PEAK)	AVERAGE ANNUAL GROWTH RATE OF REAL GNP	AVERAGE ANNUAL CHANGE IN GNP IMPLICIT DEFLATOR
1910–13	4.4%	1.4%
1913–18	2.5	11.7
1918–20	1.8	6.8
1920–23	5.4	−6.4
1923–26	3.9	0.2
1926–29	2.8	−0.5
1929–37	0.1	−1.9
1937–44	7.1	4.7
1944–48	0.1	6.7
1948–53	4.4	2.1
1953–57	2.8	2.1
1957–60	2.7	1.5
1960–69	4.3	2.6
1969–73	3.4	5.1

SOURCE: 1910–60 from J. W. Kendrick, "Concepts and Measures of Economic Growth," Table IV-1, pp. 268–69, in *Inflation, Growth, and Employment*, in the Commission on Money and Credit series, Prentice-Hall, 1964; 1960–73 computed by author.

THE GROWTH RATE AND ENVIRONMENTAL QUALITY

From the late fifties through the mid-sixties, no economic issue received more attention than the country's growth rate. The relatively slow rate during the Eisenhower years was not infrequently described as a national disaster. A major goal of the Kennedy administration that took office in January 1961 was to "get the country moving again." As shown by the annual growth rate of 4.3 percent for the 1960–69 period, there was success in this direction. But at the same time that this success was achieved, doubt began to arise as to whether this was indeed a success in a more complete sense of that word. By the late sixties there was serious questioning of a proposition that had long been practically taken for granted—that the national well-being is enhanced by an increase in the rate of output of the goods and services designed to satisfy the needs of ultimate consumers currently and in the future. What was by then being asked is whether there is a "tradeoff" between the rate at which we expand the output of goods and services and the rate at which the quality of the environment deteriorates. A more rapid rate of growth means more of the goods and services that most people want, but it also means more of the air, water, land, and noise pollution and environmental damage that people do not want. With the awakening to the seriousness of the pollution problem in the late sixties, economic

growth, which had long been generally viewed as something close to an unmixed blessing, came to be viewed by the more extreme environmentalists as something close to an unmixed evil.[9]

It is, of course, true that pollution did not suddenly appear in the late sixties. There are reports of pollution problems in London as early as 1285 as a result of the burning of soft coal. What did appear for the first time in the late sixties was the widespread awareness of the fact that pollution had reached such proportions that, if allowed to grow completely unchecked, it could destroy civilization in the course of not too many more generations. Accompanying real GNP (valued in 1972 dollars) of $533.5 billion in 1950 was a level of pollution that was apparently close to the assimilative capacity of the natural environment and thus gave rise to no great concern. However, eighteen years later, in 1968, real GNP had doubled to $1051.8 billion and with it probably came more than a proportional increase in the amount of lead and mercury and other poisons deposited annually in the air and water in the amount of environmental damage of practically every other kind. A growth rate of 4 percent will produce another doubling of real GNP by 1986. Because we have already passed the tolerable levels for certain kinds of pollution, the prospect of a doubling of existing levels understandably strikes fear into almost everyone. It is easy to understand the opposition to growth voiced by those who believe there is inevitably a close relation between the rate of growth of GNP and the rate of decay of the environment.[10]

Whether or not growth is indeed at the heart of the problem is a question we will turn to in a moment. However, even if it is, it should be seen that a cessation of growth will not provide a solution to the problem. To drop all the way down to a zero rate of economic growth would not stop the deterioration of the environment; it would only slow the rate at which it deteriorates. A zero rate of growth means an unchanged total of goods and services produced

[9]In order to evaluate the cost of economic growth in terms of environmental damage, it is necessary to somehow define each of the concepts in quantitative dimensions that permit an analysis of their relationship. Defining economic growth is relatively easy compared with defining environmental deterioration. First, the latter concept is in itself much more elusive than that of economic growth, and, second, the problem of evaluating those changes in the ecosystem that involve environmental deterioration in a way that can be related, quantitatively if possible, to the process of economic growth is extremely difficult.

[10]That there is such a close relation was the belief of the membership of the Club of Rome which sponsored the research set forth in the widely publicized bombshell of a book, *The Limits to Growth*, Universe Books, 1972, by D. H. Meadows, D. L. Meadows, et al. The authors constructed an elaborate computer model of the world that involves five basic quantities: capital stock, population, food production, nonrenewable resources, and pollution. Among the relationships in the model are complex "feedback loops" that register the effects of changes in one variable such as food production on another such as birth rates. Into the model went past data on such things as rates of growth of population, industrial output, and agricultural yields, and estimates of rates of technological advance that would allow the use of new resources, raise agricultural productivity, and control pollution. The primary conclusion that the authors reach is that unchecked growth can have only one outcome: a rather sudden and uncontrollable collapse in both population and industrial capacity some time before the year 2100. No matter which way one turns, the conclusion is always the same: collapse from shortage of food, from exhaustion of resources, or from pollution. The only possible escape is to stop the rate of growth of output and of population within the next fifteen years and achieve what is called a "global equilibrium" with zero growth. But with the likelihood that this will happen not being very great, what seems to await mankind in the next century is doomsday. The basis on which this apocalyptic vision was reached was subjected to a well-deserved barrage of criticism from prominent economists and ecologists. None of the critics denies the seriousness of the problem, but they disagree with the forecast of almost inevitable disaster. Whatever the scientific shortcomings of *The Limits to Growth*, the study had a great impact and provided a scientific basis of sorts for an increase in the antigrowth sentiment that had already been rapidly developing over the preceding several years. However, this antigrowth sentiment has probably weakened over the years since the appearance of *The Limits to Growth*. The Second Report to the Club of Rome, a volume by M. Mesarovic and E. Pestel entitled *Mankind at the Turning Point*, E. P. Dutton, 1974, did not conclude, as did the first report, that collapse was inevitable unless growth was stopped. With the second report, the Club of Rome modified its view and declared that it could find some hope in the future and spoke optimistically of an organic growth that takes prudent account of environmental and other dangers.

each year. To the degree that the amount of pollution depends on the amount of output, the result is that the amount of pollutants dumped into the system will be no larger next year than this year. However, preventing an ever larger amount of damage from being inflicted on the environment year after year is obviously not the same as protecting the environment from all damage. The problem remains: If a positive rate will lead to doom, a zero rate will do the same but on a slower timetable.

Beyond this, even if it were true that a cessation of economic growth would mean a cessation of environmental deterioration, it is not likely that the problem could be solved in this way. Since it is a world problem, it would do little good to stop growth in the United States and other affluent countries if it continued unrestrained in the less affluent and poor countries. Because the people of some poor countries see in economic growth the only possibility of rising out of poverty, it is fanciful to believe that their governments would voluntarily adopt a no-growth policy. The very suggestion of such a policy would be interpreted by them as a conspiracy by the affluent countries to lock them into perpetual poverty. To try to meet this problem through a redistribution of wealth is another unworkable solution. Redistribution on the scale required both within and between countries could not occur without force, which would mean revolutions and wars.

If the preceding is essentially correct—that a zero-growth policy in itself will not stop the deterioration of the environment but only put it on a slower timetable and that, in any event, a zero-growth, world-wide policy is not something that could be implemented in a world of nations that differ so widely in wealth—the answer to the problem of the deterioration of the environment is not to be found in stopping economic growth. However, if continuing growth means continuing environmental decay, there would seem to be no escape from eventual disaster. This, at least, would be the inevitable result if economic growth over the years ahead were of the same kind as that over the years since World War II. On the other hand, if it is possible to bring about appropriate changes in the composition of the growing output and in the technology employed to produce that output, growth in the future may do more to slow the deterioration in the environment than a situation of no growth could, assuming that the latter could be realized in practice. It may even be possible that, within limits, the more rapid the rate of growth, the better the job that can be done to slow the deterioration of the environment. If pushed to this extreme, this argument stands the argument of the zero-growth proponents on its head.

It is generally conceded that the know-how exists to provide procedures and produce systems capable of greatly reducing most kinds of pollution.[11] The particular way or ways in which the reallocation of resources needed to provide the required pollution-control facilities may be brought about is another of the many difficult questions faced in this area. Or perhaps even more basic is the question of whether or not it can even be brought about in a capitalistic system in which the profit incentive is so important

[11]For example, in the case of pollutants emitted into the air, the ones of central concern are carbon monoxide, hydrocarbons, sulfur dioxide, oxides of nitrogen, and particulates. In the United States, the greatest tonnage of these pollutants comes from the transportation sector, and almost all of this from the interal combustion engine. Some of these may be reduced by engine changes to achieve more complete combustion of fuel; others, like oxides of nitrogen, that are not a result of incomplete combustion may be controlled by the installation of catalytic afterburners, which add substantially to the complexity and cost of engines. Another possibility is a changeover to the Wankel engine, which as of 1971 met the government's pollution standards set for 1975. Beyond this is the complete abandonment of the internal combustion engine in favor of steam or electric engines if and when either of the latter is made practical. Still another alternative is greater reliance on mass transit. Control of airborne residuals from stationary sources (utility power, industry, and households) involves approaches such as (a) fuel preparation, e.g., removal of sulfur-bearing pyrites from coal before burning; (b) redesign of burners to reduce oxides of nitrogen; and (c) treatment of stack gases, e.g., scrubbing with water, to remove sulfur and particulates. See A. V. Kneese, "Background for the Economic Analysis of Environmental Pollution," in *Swedish Journal of Economics*, March 1971, pp. 1–24, reprinted in P. Bohm and A. V. Kneese, eds., *The Economics of Environment*, Macmillan, 1971.

in deciding what shall and what shall not be produced. However, assuming for the moment that society can effect a shift in the composition of output such that a substantially larger portion of total output is made up of investment in antipollution devices and facilities of various kinds, a large step would have been taken in the right direction. Systems would be acquired for recycling, which reduces pollution by channeling wastes of the production and consumption processes back into these processes. Also, systems would be acquired to augment the natural environment's assimilative capacity through such means as stream aeration and chemical waste treatment.[12]

An idea of the magnitude of the required investment in facilities and the costs of operating them is provided by the following estimates of the Council on Environmental Quality. To meet the standards established by federal environmental legislation, the estimated amounts (in billions of 1975 dollars) for both operation and maintenance costs and capital costs for the decade 1975–84 are as follows: air, $174.9; water, $248.3; solid waste, $59.6; and noise, $3.4; or a total of $486.2 or approximately one-half of a trillion dollars for the decade.[13]

Devotion of more of the nation's resources to producing output whose sole purpose is to control the amount of pollution generated by the production and use of other output can be effected more easily under conditions of a positive growth rate. For although there will be virtually universal agreement that there must be greater investment in output devoted to the control of pollution, there will be far less agreement on how the real cost of this control shall be apportioned among various sectors of society. This is why some economists argue that a growth rate well above zero is, at least in this regard, preferable to one closer to zero. With a higher growth rate, the amount of other goods that must be given up is less, and therefore it will be easier to bring about smoothly the necessary shift in the composition of output.

While a sizable shift of this kind in the composition of output is essential to meet the problem of environmental deterioration, it is not the only shift required. The composition of output should also be altered to move us back toward the way things were in the pre-World War II economy. A major factor in explaining the environmental damage over the years since World War II has been the change in the composition of output from natural products to synthetic ones (soaps to detergents, paper to plastic), of power-conservative products to relatively power-consumptive products (railroads to automobiles), and of reusable containers to "disposable" ones. These and other changes of this kind have been part of the postwar technological transformation of the economy, and according to the prominent ecologist, Barry Commoner, in most of these changes that have been part of our economic growth since 1946, the new technology has an appreciably greater damaging impact on the environment than the technology it displaced.[14] On the basis of his studies, he concludes that the postwar technological transformation of productive activities is the chief reason for the present environmental crisis.[15]

[12]Of course, the production of the aluminum, copper, steel, asbestos, and beryllium components for things like air pollution control systems and sewage plants contributes to the problem it seeks to solve. The pollution contributed by the production and operation of huge amounts of pollution control systems may thus be substantial, but still it will be only a fraction of the amount of the present pollution that it eliminates.

[13]Figures are from Table I-37, p. 167, of *Environmental Quality—The Seventh Annual Report of the Council on Environmental Quality*, September 1976.

[14]See B. Commoner, "The Environmental Costs of Economic Growth," in R. Dorfman and N. S. Dorfman, eds., *Economics of the Environment*, Norton, 2nd ed., 1977, pp. 331–53. The discussion here closely follows this article.

[15]Although the change in the product is plainly seen in a case like the change in the automobile, the technological change does not in all cases involve an appreciable change in the product itself. This is especially true in agriculture. For example, the beef from a range-fed steer is not readily distinguishable from beef from a feedlot steer. However, range-fed cattle are integrated into the soil ecosystem and produce almost zero environmental damage, whereas cattle maintained in feedlots have a considerable damaging effect.

The argument that productive activities with a large damaging effect on the environment have displaced those with much less serious damaging effect does not mean that technology is by its very nature detrimental to the environment. It does not mean that the advantages that accompany technology have to be sacrificed. What must be done is to develop new technologies that incorporate ecological wisdom as well as scientific prowess. This calls for restructuring many of the technological transformations that have occurred since 1946 in order to bring the nation's technology into balance with the undeniable demands of the ecosystem. In Professor Commoner's words:

> This will require the development of massive new technologies, including systems to return sewage and garbage directly to the soil; the replacement of synthetic materials by natural ones; the reversal of the present trend to retire soil from agriculture and to elevate the yield per acre; the development of land transport that operates with maximal fuel efficiency at low combustion temperatures; the sharp curtailment of the use of biologically active synthetic organic agents. In effect, what is required is a new period of technological transformation of the economy which will reverse the counter-ecological trends developed since 1946.[16]

The cost of this new technological transformation will run into hundreds of billions of dollars. If this transformation is undertaken, as it must be if the environment is to be saved, it will obviously have far-reaching effects on the economy. Unlike the technological transformation that occurred during the quarter century following World War II, this new one will not be undertaken in response to the profit motive. Large parts of it will occur only because it is forced through governmental regulations. It will unquestionably increase the cost of many goods and probably price some completely out of the market. Although it will bring handsome profits to those firms that win out in the competition to produce the most effective antipollution systems, it may result in shrinking profit margins in much of the rest of the business sector. This in turn could lead to a slowing of the traditional kinds of investment spending and introduce an era in which the growth rate is much slower than it was during the quarter century following World War II. It is possible that in this way a movement toward the objective of the no-growth proponents will occur automatically. Still, for reasons discussed earlier, it is to be hoped that such a movement, if it does occur, does not carry the growth rate close to zero. There is a much better chance of solving the many problems involved in carrying out a very costly full-scale technological transformation under conditions of economic growth than without growth. It appears that economic growth, which has long been recognized as a major objective of every economy, does not cease to be such an objective because of the pollution problem. Economic growth and improvement in environmental quality are not only not necessarily inconsistent but may be mutually reinforcing.

THE SOURCES OF ECONOMIC GROWTH IN THE UNITED STATES

Whether economic growth and improved environmental quality are inherently inconsistent in today's world is a question over which experts will probably continue to disagree. As we saw in the preceding section, according to one school of thought, the answer is such that they would put an end to growth forthwith. However, as a practical matter, it is likely that growth will continue but that output will be indirectly subjected to more and more controls over both the

[16] *Ibid.*, p. 353.

kinds of goods and services that can be produced and the ways in which they may be produced. Under the latter heading, the question of how to produce the ever-growing amount of energy demanded was being much debated during the seventies. As oil and natural gas reserves are depleted, what shall be the role of nuclear power, clean but according to some dangerous, coal, not so clean but safe, and other sources?

As long as growth continues, the many aspects of the complex process of growth will continue to be of great importance to economists. One of these aspects to be touched on in this section is the relative importance of the various sources of economic growth. In one classification, these sources are the rates of growth in the supply of labor, in the stock of capital, and in the productivity of these factors, this productivity being primarily dependent on the pace of technological progress. Historically, the more rapidly the quantity of labor and the quantity of capital have increased and the more rapidly technology has advanced in any time period, the greater has been the growth rate recorded for that time period. Of course, in any period where there has been serious underutilization of a growing capacity to produce, the growth rate has been adversely affected.

Attempts have been made to allocate the realized rates of growth of specific time periods to these sources (and also to the subsources into which each may be divided) and to measure the relative importance of each in the growth actually realized by the economy. From these attempts to dissect the growth record of the past come clues to what the growth record of the future may be. If we accept, for example, the finding that for the four decades from 1929 to 1969 the expanded amount of labor accounted for something more than one-third of the growth of output, the enlarged stock of capital for something more than one-eighth, and increased productivity for the remainder, estimates of the changes expected in these sources in the future will give us at least some idea of what to expect.

The most ambitious dissection of the growth rate for the U.S. economy is that carried out by Edward F. Denison.[17] Table 20-3 provides some summary data for the 1929–69 period and for the 1929–48 and 1948–69 periods into which the overall period has been divided. In the first set of three columns, we see that the growth rate of national income is 3.41 percent for the overall period, 2.75 percent for 1929–48, and 4.02 percent for 1948–69.[18] The column of figures for each of these three growth rates

[17]E. F. Denison, *Accounting for United States Economic Growth 1929–1969*, The Brookings Institution, 1974. In an earlier study *The Sources of Economic Growth in the United States*, Supplementary Paper No. 13, Committee for Economic Development, 1962, Denison published a similar set of estimates for the United States for 1909–57. He also applied the viewpoint and methodology developed in the original study to the 1950–62 experience of the United States and eight western European countries in *Why Growth Rates Differ*, Brookings Institution, 1967. In both of the studies for the United States, he identifies a large number of sources of growth (nineteen in Table 20-3) and estimates the portion of the growth rate attributable to each. In earlier studies, other economists worked with only three or four, their major objective commonly being to identify the portion of growth due to changes in output per unit of input and the portion due to changes in the total amount of input. Among these studies are M. Abramovitz, *Resources and Output Trends in the United States since 1870*, National Bureau of Economic Research, 1959; F. C. Mills, *Productivity and Economic Progress*, Occasional Paper 38, National Bureau of Economic Research, 1952; J. Schmookler, "The Changing Efficiency of the American Economy, 1869–1938," in *Review of Economics and Statistics*, Aug. 1952, pp. 214–31; J. W. Kendrick, *Productivity Trends in the United States*, National Bureau of Economic Research, 1961; and R. M. Solow, "Technical Progress, Capital Formation, and Economic Growth," in *American Economic Review*, May 1962, pp. 76–86.

[18]Note that the output measure here is national income rather than the more familiar GNP. Note also that national income is potential rather than actual. Based on actual national income, the growth rates for the indicated periods are 3.33, 2.75, and 3.85 percent respectively. A difference between the growth rate for actual national income and potential national income in any period reflects the fact that the first and last years of that period are not comparable in terms of business cycle phase or of short-term demand pressures. Although 1929, 1948, and 1969 are all peak years of business cycle expansions, demand pressure was nonetheless less strong in 1969 than in the other two years. Therefore, increases in actual output from 1929 to 1969 and from 1948 to 1969 were less than increases in potential output. To secure growth estimates unaffected by the short-term status of demand in the beginning and

TABLE 20-3
Allocation of growth rate of national income among the sources of growth

	PERCENTAGE POINTS IN GROWTH RATE			PERCENT OF GROWTH RATE		
	1929–69	*1929–48*	*1948–69*	*1929–69*	*1929–48*	*1948–69*
National income	3.41	2.75	4.02	100.0	100.0	100.0
Total factor input	1.82	1.50	2.11	53.4	54.5	52.5
Labor	1.32	1.37	1.31	38.7	49.8	32.6
Employment	1.09	1.05	1.15	32.0	38.2	28.6
Hours	−0.22	−0.23	−0.19	−6.5	−8.4	−4.7
Average hours	−0.49	−0.64	−0.34	−14.4	−23.3	−8.5
Efficiency offset	0.19	0.33	0.06	5.6	12.0	1.5
Intergroup shift offset	0.08	0.08	0.09	2.3	2.9	2.2
Age-sex composition	−0.05	0.00	−0.10	−1.5	0.0	−2.5
Education	0.41	0.39	0.42	12.0	14.2	10.4
Unallocated	0.09	0.16	0.03	2.6	5.8	0.7
Capital	0.50	0.13	0.80	14.7	4.7	19.9
Inventories	0.09	0.05	0.12	2.6	1.8	3.0
Nonresidential structures and equipment	0.20	0.03	0.36	5.9	1.1	9.0
Dwellings	0.19	0.06	0.29	5.6	2.2	7.2
International assets	0.02	−0.01	0.03	0.6	−0.4	0.7
Land	0.00	0.00	0.00	0.0	0.0	0.0
Output per unit of input	1.59	1.25	1.91	46.6	45.5	47.5
Advances in knowledge and n.e.c.[a]	0.92	0.62	1.19	27.0	22.5	29.6
Improved resource allocation	0.30	0.31	0.31	8.8	11.3	7.7
Farm	0.26	0.28	0.24	7.6	10.2	6.0
Nonfarm self-employment	0.04	0.03	0.07	1.2	1.1	1.7
Dwellings occupancy ratio	0.01	0.02	−0.01	0.3	0.7	−0.2
Economies of scale	0.36	0.29	0.43	10.6	10.5	10.7
Irregular factors	0.00	0.01	−0.01	0.0	0.4	−0.2
Weather in farming	0.00	0.01	−0.01	0.0	0.4	−0.2
Labor disputes	0.00	0.00	0.00	0.0	0.0	0.0

[a]n.e.c. Not elsewhere classified.
SOURCE: E. F. Denison, *Accounting for United States Economic Growth, 1929–1969*, The Brookings Institution, 1974, pp. 127–28.

terminal years of the periods chosen, Denison estimated potential national income for the years bounding the periods and calculated growth rates on the basis of these figures. It was pointed out on p. 386 that growth rates of actual output for periods which start or end at different phases of the business cycle are not strictly comparable. We see here that one approach to this problem is to replace the figures for actual output with figures for potential output as Denison has done. See *Accounting for United States Economic Growth, 1929–1969*, pp. 9, 127–28.

shows the estimated portion of each accounted for by each of a number of sources to which that growth rate can be attributed. In broad terms, we find that of the 3.41 percent growth rate for 1929–69, 1.82 percentage points are accounted for by total factor input and 1.59 percentage points by output per unit of input (or productivity). Of the 1.82 percentage points, 1.32 are accounted for by labor and 0.50 by capital. Further breakdowns are given for each of these input headings as well as for output per unit of input.

If one looks at the breakdowns for the 1929–48 and 1948–69 periods, capital and advances in knowledge are seen to be primarily responsible for the marked rise of 1.27 percentage points in the growth rate from 2.75 percent in 1929–48 to 4.02 percent in 1948–69. From the earlier to the later period, capital increased its contribution by 0.67 percentage points and advances in knowledge increased its contribution by 0.57 percentage points. Denison observes that the latter change is probably due largely to a faster rate of advances in knowledge relevant to production in the later period but that an end to the depression-induced restrictions against efficient practices also contributed to it. Examination of the figures shows that the 0.14 percentage point increase from 0.29 in 1929–48 to 0.43 in 1948–69 in the growth rate that is attributed to economies of scale makes this source next in importance in accounting for the rise in the growth rate between the two periods. The increase assigned to economies of scale was the consequence of the faster expansion of markets, but it was in turn the changes in the contribution of capital and advances in knowledge that made the faster expansion of markets possible.

The second set of three columns converts the first set into percentages that give the percent of the growth rate (e.g., 3.41 = 100) accounted for by each of the sources and thereby permits a direct comparison of the relative importance of each of the sources in accounting for the growth rate in any period and of the relative importance of each of the sources in different periods. Thus, the first of these columns shows that for the 1929–69 period total factor input—increasing the quantity of labor and capital—was responsible for 53.4 percent of total growth and that productivity—getting more output per unit of input—was responsible for 46.6 percent. Of the inputs, labor was responsible for 38.7 percent of total growth and capital for only 14.7 percent. The 1929–48 and 1948–69 columns reveal major changes in the percentages. The percentage for capital was atypically small in 1929–48, a period during which it accounted for only 4.7 percent of total growth in comparison with 19.9 for 1948–69. As the percentages accounted for by total factor input in 1929–48 and 1948–69 are similar, 54.5 and 52.5 percent respectively, the unusually low percentage contribution for capital in the 1929–48 period gives an unusually high percentage contribution for labor in the same period.

A look at the figures for the subheadings reveals the interesting conclusion that advances in knowledge—technological, managerial, and organizational—contributed a little more to the growth rate than the increase in the amount of work done by the labor force for the 1929–69 period and contributed substantially more for the 1948–69 period. A major part of the contribution made by labor is not measured by the amount of work done but by the education of the labor force. Taking the sum of the percentages for employment and hours as the percentage for work done, we find that for 1929–69 the percentage for work done is 25.5 or somewhat less than the percentage of 27.0 for advances in knowledge. If the same is done for the 1948–69 period alone, the result is that the 29.6 percentage attributed to advances in knowledge exceeds by far the 23.9 percent for employment and hours or work done. When Denison's estimates are classified in this way, advances in knowledge emerge as the most important source of growth.

However, if one views the contribution of labor not just in terms of the amount of work

done by employed persons but also allows for education of employed persons, the contribution of labor becomes that shown in the table under the heading "labor" and it emerges as the most important source of growth in all three periods, especially in the 1929–48 period during which it accounts for almost one-half of the growth rate. From this view, advances in knowledge then takes second place. Third in importance for the 1929–69 and the 1948–69 periods is the growth in the country's capital stock. Fourth comes economies of scale, which largely reflect the expansion of markets. And next is improved resource allocation, which for the most part is the correction of the misallocation of labor to farming and to small and inefficient family enterprises in the nonfarm sector. As the other sources are of negligible importance, it may be concluded that the five sources listed account for practically all of the growth of the U.S. economy over the forty-year period, according to the estimates made by Denison.

Although Denison has provided the most detailed allocation of the growth rate of the U.S. economy among its various subsources, his basic conclusion on the initial allocation between total factor input and output per unit of input is in line with that reached in other, earlier studies. However output per unit of input may be allocated into the parts attributable to advance of knowledge, economies of scale, and other subsources, the output per unit of input in itself accounts for close to half of the growth rate. Although different studies assign somewhat different importance to increases in productivity, they all reach the conclusion that, in the words of Moses Abramovitz, "to explain a very large part of the growth of total output and the great bulk of output *per capita*, we must explain the increase in output per unit of conventionally measured inputs. . . ."[19]

ECONOMIC GROWTH—DEVELOPED AND DEVELOPING ECONOMIES

The problems of growth are not limited to developed capitalistic economies. In the area of pollution, the Volga in communist Russia is just as polluted as some of our major rivers, and pollution runs in the streets in many developing countries. As there are similarities, so are there differences. The sources of economic growth are quite uniform for developed countries, and a quantitative study of these sources for any developed country could proceed along much the same lines as those followed by Denison in his studies of the U.S. and western European economies. Although the relative quantitative importance of each of the primary sources of growth will not be the same in all developed economies, the primary sources themselves will be the same. In the same way, a theory of economic growth that attempts to explain the process by which growth takes place in a developed economy will, in general, be applicable to any developed economy. However, the same growth theory will not be applicable to a developing economy. Since the next chapter is devoted entirely to growth theory relevant to developed economies, it may be helpful at this point to mention the pertinent distinctions between growth in the two types of economies.

To begin with, the most serviceable and most commonly used criterion for classifying economies as developed or developing is per capita real income.[20] Those with "low" per capita real income are considered to be developing; those

[19]M. Abramovitz, "Economic Growth in the United States," *American Economic Review*, Sept. 1962, p. 776.
[20]For a discussion of this and other criteria, see B. Higgins, *Economic Development*, rev. ed., Norton, 1968, Ch. 1.

with "high" per capita real income are developed. Income is judged to be high or low by comparing it to income in countries such as Switzerland and Sweden, which apart from the oil bloc countries are the "highest," or to the United States, Denmark, and Canada, which are relatively "high." It is the level of absolute per capita real income in one country relative to that in another that is relevant for this classification. The parallel criterion for classifying economies according to their growth rates for any particular period of years is the rate at which per capita real income grows. Notice that although developing countries are typically countries with low growth rates, it is possible for such a country to show a high growth rate. If sustained, a high growth rate will clearly move the country out of the "developing" class.

If our purpose is only to compare how developed and developing economies have performed in terms of growth, we have a single problem whose solution could be provided by comparable data on per capita real income over time for both kinds of economies. However, if our purpose is to explain the process by which economic growth takes place in developed and developing economies, we are faced with two reasonably distinct problems. Because the two economies differ, economic growth theory has developed in the postwar period along two branches, one—called *development economics*—that attempts to explain the growth process in economies moving toward the developed stage, and another that attempts to explain the growth process in economies that are already developed. The two branches overlap to some extent, but there are still essential differences that make the separation analytically useful.

Developing economies are typically characterized by insufficient capital equipment, technological backwardness, structural unemployment, extreme income disparities, foreign indebtedness, low educational attainment, an unskilled labor force, and, in many cases, a deficiency of natural resources relative to the size of the population. It is essentially these characteristics that make for the relatively low per capita real income that results in the designation "developing." Therefore, to explain the slow rate (or complete absence) of growth in such economies, we must start with the explanation of the social, cultural, and economic factors that give rise to certain of these characteristics. The explanation of growth in developed economies, in contrast, takes *as given* the existence of a more or less adequate capital stock, a skilled labor force and an experienced managerial group, and technological progressiveness, and it seeks to explain the growth process in terms of the advances in output made possible by the expansion in the quantity of these basic determinants and the improvement of their quality.

What is thus taken as given for the developed economies is absent to a greater or lesser degree in the developing economies. To move up to the highroad of economic progress, developing economies must typically undergo an economic and social reorganization, and in some cases even a transformation of the whole system, involving the adoption of more advanced methods of production despite the cultural barriers to such steps, the development of efficient transportation and communications systems, the establishment of schools and facilities for public services (often termed *social capital*), and the development of modern monetary machinery that includes a capital market and other credit facilities. In some cases, these economies must first put an end to political instability and produce a climate conducive to the attraction of foreign capital. Broad changes are often needed in the economy's product "mix," and the increased mobility and relocation of labor needed to support such changes must be secured. Finally, there must be mass education and the provision of the necessary industrial skills to adapt labor to new types of work. These and many other developments must proceed more or less simultaneously. Included among these other developments are

the accumulation of capital, the expansion of the quantity and quality of the labor force, and the achievement of the technological advances that are recognized as the basic determinants of growth in an already developed economy.

Since these and other developments are all typically involved in the problem of growth in a developing economy, much deeper "digging" is clearly called for in any attempt to explain the growth process in such economies. The theory of growth for a developed economy thus turns out to be the less complex of the two. Chapter 21 is devoted entirely to growth theory of this simpler kind. Such theory is not completely without relevance to the developing economies, but it is still essential to note that there are important underlying differences in the theory of economic growth in the two cases.

chapter 21

Economic Growth Theory

Keynes's *General Theory* not only began a new chapter in the short-run theory of income and employment but produced as a by-product a revival of interest in long-run theory, or the theory of economic growth. In view of the impact made by the *General Theory*, it was only to be expected that economists would proceed to construct theories, which would secularize and make dynamic the short-run static theory presented by Keynes. Since the late thirties a huge body of growth theory has developed that not only has extended Keynesian theory to the long run but has gone far beyond this. Thus, Professors Hahn and Matthews, in presenting what is only a survey of the theory of economic growth for the period from the *General Theory* to the early sixties, require more than one hundred pages to complete the task.[1]

In this chapter we will touch on only two but probably the two most basic growth theories that have been developed since the thirties. The theory that serves as the point of departure in the survey mentioned above and the one that usually receives first consideration in any current discussion of modern growth theory is one that relates the growth rate of the economy's aggregate output to that of its capital stock. In this approach, capital is the only factor of production explicitly considered, and it is assumed that labor is combined with capital in fixed proportions. With regard to the rate at which capital accumulates, this theory is Keynesian in nature. It holds that under laissez-faire conditions there is no effective adjustment mechanism to equate investment with full-employment saving. If it is true that changes in thriftiness do not automatically lead to equal changes in investment, the rate of growth will not be that which would occur with capital accumulation determined by full-employment saving. This fixed-proportion, Keynesian-based growth theory, the first to be examined in this chapter, is most commonly known as the *Harrod-Domar theory*.

So rapid and so varied has been the development of growth theory in the last few decades that the Harrod-Domar theory may now be described as a relatively early and relatively simple approach to the growth question. Most of the growth theory that followed and the theory that has dominated the area in the past twenty years is known as *neoclassical growth theory*. In this theory, the second one we will look into in this chapter, a production function is employed in which capital and labor may be combined in varying proportions. Unlike the Harrod-Domar

[1] F. H. Hahn and R. C. O. Matthews, "The Theory of Economic Growth: A Survey," in *Economic Journal*, Dec. 1964, pp. 779–902.

model, substitutability among factors is thus allowed for in the production process. Also unlike the other theory, the neoclassical theory assumes that capital will accumulate at a rate set by the thriftiness of the economy operating at full employment. The question of whether or not the amount of saving forthcoming at full employment will be matched by an equal amount of planned investment, a prerequisite to the continued full utilization of the factors of production, is thus answered in terms of classical and not Keynesian theory.

There is a third branch of recent growth theory that is commonly described as the modern Cambridge theory because of its close association with the names of Joan Robinson and Nicholas Kaldor of Cambridge University. It is theory that is highly critical of the neoclassical theory in many ways. For example, it rejects that theory's classical approach to saving and investment and returns to the Keynesian approach in which investment is determined not by saving propensities but by business persons' decisions, which are held to depend on such things as their experiences in the recent past, governmental policies, and sociocultural influences on the willingness to bear risk. It also rejects the neoclassical production function and with it the entire marginal productivity theory of income distribution, offering alternative macroeconomic theories as substitutes. The theory of the Cambridge school is thus both a theory of income distribution and a theory of growth in which one of the unique features of the growth theory is its incorporation of a "Keynesian" theory of income distribution. Although the attack of Robinson and Kaldor and others on the neoclassical growth theory and the defense thereof by its supporters have filled the journals with dozens of articles over the years since 1960, space does not permit us a more detailed examination of the modern Cambridge theory.[2]

While it is true that the last thirty-five years have been a period of unprecedentedly rapid development of growth theory, it must be noted that economists' concern with the question of economic growth goes back at least as far as 1776, to Adam Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations*. The very title of this classic suggests the author's interest in the long-run question of the accumulation of wealth, one aspect of any study of economic growth. In Smith's view of the process by which the economy's wealth expands, an important role is assigned to the "division of labor," or specialization. The idea of diminishing returns had not yet been "discovered," and Smith saw no obstacle to the increase of returns to labor through increased specialization. His conclusion was thus an optimistic one in which increased specialization would lead to a rising standard of living for a growing population. This optimistic conclusion was replaced a few decades later by a pessimistic one as a result of the work of Thomas Malthus and David Ricardo: The Malthusian population principle and the Ricardian diminishing returns and rent theory combine to force the great majority of the economy's population down to a subsistence standard of living. Between the writings of classical economists such as these and the writings of the economists of the past thirty-five years are such major contributions as the growth systems of Karl Marx and Joseph Schumpeter, both of whom saw the capitalistic process as one in which business persons are engaged in an unrelenting drive toward accumulation that would one day end with the destruction of the system.[3] In short, growth theory is a product not of the past thirty-five years but rather of the past two hundred years.

[2]For an analysis of the modern Cambridge models, see D. Hamberg, *Models of Economic Growth*, Harper & Row, 1971, Ch. 3, and H. Y. Wan, Jr., *Economic Growth*, Harcourt Brace Jovanovich, 1971, Ch. 3. Further references will be found in these chapters.

[3]For an analysis of classical growth theory and the growth theory of Marx and Schumpeter, see, for example, B. Higgins, *Economic Development*, rev. ed., Norton, 1968, Chs. 3–5; W. J. Baumol, *Economic Dynamics*, 3rd ed., Macmillan, 1970, Chs. 2–3; and D. Hamberg, *Economic Growth and Instability*, Norton, 1956, Ch. 1.

HARROD-DOMAR GROWTH THEORY[4]

A basic principle emphasized by Harrod and Domar and incorporated in all modern growth theory is the dual effect of net investment: Net investment constitutes a demand for output, and it also increases the capacity of the economy to produce output. For example, constructing and equipping a new factory generates a demand for steel, brick, and machinery, and the factory once constructed and equipped increases the economy's productive capacity. The economy's net investment in any period thus has a *demand* and a *capacity* effect. If the amount of net investment in any period equals that period's net saving, it has the demand effect of removing from the market that portion of output not purchased by consumers, thereby making aggregate demand equal to aggregate output for the period and making that period's actual level of income and output the equilibrium level. This is nothing more than the familiar Keynesian principle that there must be equality between planned saving and planned investment if there is to be equilibrium in the level of income and output. What is not familiar from the Keynesian model is the fact that this period's net investment also has a capacity effect: It increases the economy's productive capacity in this period and thereby increases the next period's potential output. If this expanded capacity is to be fully utilized, aggregate demand in the next period will have to exceed that of this period. Thus, in general, as long as there is net investment in one period after another, aggregate demand must rise period after period if the expanding productive capacity resulting from net investment is to be fully utilized.

Increase in Capacity Output

The basic theory involves a simple production function that relates the generation of total output to the stock of capital via the capital–output ratio. Taking the techniques of production as given, some specified amount of capital goods is necessary to produce a given amount of output. If we let K represent the capital stock and Y the level of output, we may define the average capital–output ratio as K/Y. Thus, if it takes \$3 worth of capital goods to produce \$1 worth of final output, a capital stock of \$300 billion is required to produce aggregate final output of \$100 billion, and the capital–output ratio is 3 to 1.[5] In contrast, the marginal capital–output ratio $\Delta K/\Delta Y$ tells us how much additional capital is necessary to produce a specified addition to that flow of output. The marginal ratio need not be equal to the average

[4]The basic elements of the Domar model are found in Evsey Domar, "Expansion and Employment," in *American Economic Review*, March 1947, pp. 34–55. See also his "Capital Expansion, Rate of Growth, and Employment," in *Econometrica*, April 1946, pp. 137–47, and "The Problem of Capital Accumulation," in *American Economic Review*, Dec. 1948, pp. 777–94. These and related essays by Domar are reprinted in his *Essays in the Theory of Economic Growth*, Oxford Univ. Press, 1957. For the Harrod model, see Roy F. Harrod, "An Essay in Dynamic Theory," in *Economic Journal*, March 1939, pp. 14–33, reprinted in A. H. Hansen and R. V. Clemence, eds., *Readings in Business Cycles and National Income*, Norton, 1953, pp. 200–19, and revised and expanded in Harrod's *Towards a Dynamic Economics*, St. Martin's, 1948 (see particularly pp. 63–100). The 1939 essay is refined but unchanged in essentials in Harrod's "Second Essay in Dynamic Theory," in *Economic Journal*, June 1960, pp. 277–93. On both models, see also D. Hamberg, *Models of Economic Growth*, Ch. 1; and H. Y. Wan, Jr., *Economic Growth*, Ch. 1.

[5]Note that the ratio depends on the time period chosen for the measurement of output. Capital stock of \$300 billion will produce, say, \$100 billion of output *per year*—that is, the 3 to 1 ratio. The same capital stock will produce \$25 billion of output *per quarter*, a 12 to 1 ratio, and \$8.3 billion *per month*, a 36 to 1 ratio. One year is the time period usually selected for measurement of flows. Note also that our analysis makes use of the average capital–output ratio for the economy as a whole; the capital–output ratio will, of course, vary among the different goods into which aggregate output may be broken down.

ratio as long as technology changes over time. It would rise with capital-using technological changes and fall with capital-saving technological changes. In the model, however, we assume a constant technology, so that $\Delta K/\Delta Y$ remains constant. To simplify the analysis still further, we assume that the constant $\Delta K/\Delta Y$ equals K/Y, so that K/Y is also constant.

The reciprocal of the average capital–output ratio, Y/K, represents the average productivity of capital. If \$3 of capital goods is required to produce \$1 of final output—that is, if $K/Y = 3$—the average productivity of capital, Y/K, will be 1/3, or 0.33. Given an increase in the capital stock, ΔK, $\Delta Y/\Delta K$ indicates the ratio of the increase in capacity output to the increase in capital stock. Just as $\Delta K/\Delta Y$ need not be the same as K/Y, so $\Delta Y/\Delta K$ need not be the same as Y/K. The pace of technological advance will affect these ratios, but the simple model does not treat them as variables, so that $Y/K = \Delta Y/\Delta K$. This ratio of output to capital stock is designated by σ (sigma), which Domar calls the "potential social average productivity of capital."

Since ΔK in any period equals that period's net investment, I, $\Delta Y/\Delta K = \sigma$ may also be expressed as $\Delta Y/I = \sigma$ or $\Delta Y = \sigma I$. Thus, if $\sigma = 1/3$, \$1 of net investment will increase capacity output by 33.3¢. From this it follows that the cumulative net investment of any period increases capacity output by σI. This is the most important relationship in the model. By defining

Increase in Aggregate Spending

In a two-sector economy, aggregate spending equals the sum of consumption and investment expenditures. With a stable consumption function, consumption expenditures will rise only as a result of a rise in income, and therefore a rise in investment expenditures is necessary to initiate a rise in income. Consequently, we may determine the total rise in expenditures, or income, that will result from any given rise in investment by using the simple multiplier expression

$$\Delta Y = \frac{1}{s} \Delta I$$

in which s is the MPS.[6] This rise in income, or expenditures, is matched by an equal rise in actual output, since, with a stable price level, output responds in proportion to the rise in spending. With subscript r designating realized, or actual, we have

$$\Delta Y_r = \frac{\Delta I}{s}$$

This brings us to a crucial asymmetry of the demand and capacity effects of investment. On the demand side, an increase in actual output, ΔY_r, is a function not of investment but of the *increment* to investment. (Recall from the basic Keynesian model that if net investment remains unchanged period after period, aggregate spending will remain unchanged and so will aggregate

tential output expands period by period, but actual output remains unchanged, so unused productive capacity must result. The assumed constant net investment will not be maintained by business period after period in the face of such excess capacity. This underscores the essential paradox of investment: To justify today's net investment, tomorrow's must exceed today's in order to provide the additional aggregate spending needed to purchase that part of the enlarged potential output that is not purchased by consumers. In other words, as long as net investment is positive, it must increase in order not to decrease. The economy cannot stand still period after period with a constant net investment; either it will move ahead (e.g., if autonomous consumption or government spending increases), or it will fall back. Only a stationary economy neither moves ahead nor falls back, but by definition there is zero net investment in a stationary economy. Zero net investment is not characteristic of any developed real-world economy. If, then, a real-world economy must be a growing economy, what is the rate at which it must move ahead to avoid falling back? In other words, what is its equilibrium rate of growth?

The Equilibrium Rate of Growth

There is some rate of growth at which the increase in actual output in each period, ΔY_r, will just equal that period's increase in capacity output, ΔY_p. This rate at which $\Delta Y_r = \Delta Y_p$ is called the *equilibrium rate of growth*. Of course, ΔY_r and ΔY_p are not rates as such but absolute amounts of change measured from one period to the next. However, by relating each such absolute amount of change to the absolute actual amount of output and to the absolute capacity output of the preceding period, each may be expressed as a rate. Thus, from an original period in which there is equilibrium as given by $Y_r = Y_p$, it follows that if the rate $\Delta Y_r/Y_r$ remains equal to the rate $\Delta Y_p/Y_p$ period after period, Y_r will remain equal to Y_p period after period. In such a situation, aggregate realized output grows as fast as aggregate potential output, thereby producing a path of equilibrium growth over time.

Since $\Delta Y_r = \Delta I/s$ and $\Delta Y_p = \sigma I$, the equilibrium rate is also that at which $\Delta I/s = \sigma I$. The left side shows the increment to aggregate realized output for the period; since this is equal to the increment of aggregate spending, it may be called the demand side. The right side shows the increment to productive capacity for the period and as such may be called the supply side. Note that investment appears on both sides of the equation, although not in the same form. On the left side we have the increment to net investment, or the difference between the net investment of any one period and that of the preceding. This absolute change of net investment times the multiplier determines the change in aggregate output. On the right side of the equation, however, we find not the *change* in net investment for the period but the *total* net investment for the period. The reason, of course, is that total net investment for the period times the average productivity of capital determines the change in productive capacity.

We may solve this equation by multiplying both sides by s and then dividing both sides by I. Given that

$$\frac{\Delta I}{s} = \sigma I$$

multiplying by s yields

$$\Delta I = s\sigma I$$

Then, dividing by I,

$$\frac{\Delta I}{I} = s\sigma$$

In this form the left side of the equation now gives the required rate of growth of net investment. If actual output is to rise as fast as potential output, the growth rate of net investment

must be $s\sigma$, or the propensity to save multiplied by the productivity of capital.

Although ΔI is subject to a multiplier that makes ΔY greater than ΔI, we can see that the growth rate of actual output, $\Delta Y_r/Y_r$, must be the same as the growth rate of investment, $\Delta I/I$. Since in equilibrium $\Delta Y_r = \Delta Y_p$ and since $\Delta Y_p = \sigma I$, it follows that $\Delta Y_r = \sigma I$. Furthermore, since $I = sY$ in equilibrium, then by substitution $\Delta Y = \sigma sY$ and $\Delta Y/Y = s\sigma$. Therefore,

$$\frac{\Delta I}{I} = \frac{\Delta Y}{Y} = s\sigma$$

The rate at which actual output and investment must grow in order that actual output remain equal to potential output is determined by the propensity to save and the productivity of capital. The higher the propensity to save, the greater is the required growth rate, and conversely. The higher the productivity of capital, the greater is the required growth rate, and conversely. The meaning of these relationships may be brought out most clearly through the series of numerical illustrations that follow.

Numerical Illustrations of the Growth Process

Assume an economy in which for a given year aggregate spending equals the aggregate potential output that economy can produce with the capital stock existing at the beginning of the year. This defines an equilibrium for the year (say year 1) in which $Y_r = Y_p$ and provides a take-off point from which we can trace the growth process over the next few years. Re- put, Y_p, of 100 [column (3)]. Let us also assume that in year 1 actual, or realized, output, Y_r [column (4)], was 100. Consumption [column (5)], equal to $(1 - s)Y_r$, was 80, so saving, equal to sY_r, was 20. Investment [column (7)] was also 20, yielding aggregate spending of 100, the amount necessary to make realized output equal to potential output. This much is the ordinary, short-run Keynesian model. Now allow for the fact that I of 20 in year 1 increases by 20 the capital stock in existence at the beginning of year 2. This increases capacity output of year 2 by σI (or $\sigma\Delta K$)—that is, by 5. This means that aggregate spending in year 2 must also increase by σI to make use of the increase in capacity. But aggregate spending will increase only if investment increases. To secure the required increase in aggregate spending of 5 in year 2, investment must rise from 20 to 21. Given the multiplier of 5 ($s = 1/5$), we then find an increase in spending of 5 composed of $\Delta C = 4$ [column (6)][7] and $\Delta I = 1$. If investment in year 2 does in fact rise at the required rate of 5 percent, or from 20 to 21, Y_r in year 2 will equal Y_p in year 2. Or, given that $\Delta I/I = s\sigma$, numerically for year 2 we have $1/20 = 0.20 \times 0.25$. The growth rate of output, $\Delta Y/Y = s\sigma$, is then $(4 + 1)/100 = 0.20 \times 0.25$, where 4 represents the increase in consumption and 1 represents the increase in investment. With s and σ as given in Model 1 and as long as investment rises 5 percent per year, aggregate spending will rise as fast as potential output, and actual output will grow accordingly at the capacity rate of 5 percent. At this constant r

TABLE 21-1
Illustrations of the growth process: equilibrium situations

(1) YEAR	(2) CAPITAL STOCK K	(3) POTENTIAL OUTPUT $Y_p = \sigma K$	(4) REALIZED OUTPUT $Y_r = C + I$	(5) CONSUMPTION $C = (1-s)Y_r$	(6) CHANGE IN CONSUMPTION $\Delta C = (1-s)\frac{\Delta I}{s}$	(7) INVESTMENT (AUTONOMOUS) I	(8) CHANGE IN INVESTMENT (AUTONOMOUS) ΔI
	Model 1: $s = 0.20$, $\sigma = 0.25$, and $\Delta I/I = \Delta Y/Y = 0.05$						
1	400.00	100.00	100.00	80.00	—	20.00	—
2	420.00	105.00	105.00	84.00	4.00	21.00	1.00
3	441.00	110.25	110.25	88.20	4.20	22.05	1.05
4	463.05	115.76	115.76	92.61	4.41	23.14	1.09
	Model 2: $s = 0.10$, $\sigma = 0.25$, and $\Delta I/I = \Delta Y/Y = 0.025$						
1	400.00	100.00	100.00	90.00	—	10.00	—
2	410.00	102.50	102.50	92.25	2.25	10.25	0.250
3	420.25	105.06	105.60	94.56	2.31	10.51	.256
4	430.76	107.69	107.69	96.62	2.36	10.77	.263
	Model 3: $s = 0.20$, $\sigma = 0.50$, and $\Delta I/I = \Delta Y/Y = 0.10$						
1	200.0	100.0	100.0	80.0	—	20.0	—
2	220.0	110.0	110.0	88.0	8.0	22.0	2.0
3	242.0	121.0	121.0	96.8	8.8	24.2	2.2
4	266.2	133.1	133.1	106.5	9.7	26.6	2.4

Models 2 and 3 show how different values of σ and s affect the equilibrium growth rate.[8] The choice of the particular values of σ and s in these three models brings out the following points. First, the growth rate in Model 1 is twice that in Model 2 because the propensity to save is twice as large (0.20 instead of 0.10) with the same productivity of capital in both cases. An economy that can double the fraction of its resources that is diverted from the production of consumer goods can double the fraction of its resources that is devoted to capital accumulation. This means that if the productivity of capital remains unchanged, the economy can grow *potentially* at twice its previous rate with what amounts to a doubling of its propensity to save. Next, we find that the growth rate in Model 3 is twice that in Model 1 because the productivity of capital is twice as great (0.50 instead of 0.25) with the

[8]Since output in year 1 is set at 100 in all three models, the original capital stock in Model 3 is only 200 with productivity of capital assumed to be 0.50.

TABLE 21-2
Illustrations of the growth process: disequilibrium situations

(1) YEAR	(2) CAPITAL STOCK K	(3) POTENTIAL OUTPUT $Y_p = \sigma K$	(4) REALIZED OUTPUT $Y_r = C + I$	(5) CONSUMPTION $C = (1 - s)Y_r$	(6) CHANGE IN CONSUMPTION $\Delta C = (1 - s)\frac{\Delta I}{s}$	(7) INVESTMENT (AUTONOMOUS) I	(8) CHANGE IN INVESTMENT (AUTONOMOUS) ΔI
			Model 1-A: Investment grows at the required rate of 5 percent.				
1	400.00	100.00	100.00	80.00	—	20.00	—
2	420.00	105.00	105.00	84.00	4.00	21.00	1.00
3	441.00	110.25	110.25	88.20	4.20	22.05	1.05
4	463.05	115.76	115.76	92.61	4.41	23.14	1.09
			Model 1-B: Investment is constant.				
1	400	100	100	80	—	20	—
2	420	105	100	80	—	20	—
3	440	110	100	80	—	20	—
4	460	115	100	80	—	20	—
			Model 1-C: Investment grows at the too slow rate of 3 percent.				
1	400.00	100.00	100.00	80.00	—	20.00	—
2	420.00	105.00	103.00	82.40	2.40	20.60	0.60
3	440.60	110.15	106.09	84.87	2.47	21.22	.62
4	461.82	115.45	109.28	87.42	2.55	21.86	.64

same propensity to save in both cases. This suggests that if the amount of output obtainable per unit of capital can somehow be doubled, an economy can double its *potential* rate of growth with the unchanged rate of capital accumulation set by its propensity to save. Finally, we note that the growth rate in Model 3 is four times the growth rate in Model 2 because both the productivity of capital and the propensity to save are twice what they are in Model 2.

Each of the models of Table 21-1 traces out an equilibrium growth path, the sole difference between them being in the values of σ and s, which determine the rate of equilibrium growth. In each model the economy follows an equilibrium growth path as a result of growth in aggregate spending equal to the growth in capacity output, a result that in turn depends on growth of investment at the required rate. Model 1-A of Table 21-2 repeats Model 1 of Table 21-1, in which investment grows at the required rate of 5 percent. Models 1-B and 1-C, however,

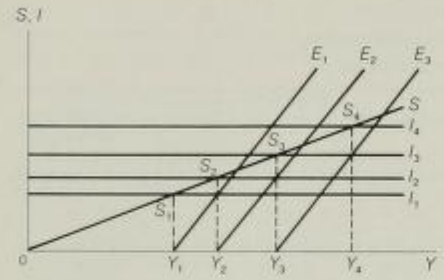


FIGURE 21-1
The equilibrium growth process

show what happens to growth if investment fails to grow at the rate required to make full use of the growing capacity. In Model 1-B, net investment remains constant, and in Model 1-C, net investment grows but at less than the required rate.

The last two models describe disequilibrium situations. In Model 1-B, in which there is a constant net investment of 20 per year, there is a constant addition to capacity of 5 each year. Since changes in aggregate spending depend on changes in net investment and not on the level of net investment, aggregate spending remains constant as long as net investment is constant. Since with each passing year 5 more is being added to excess capacity under these circumstances, it is not likely that net investment will be maintained even at 20 year after year. A decrease in net investment will follow, and this will cause aggregate spending to fall; depression and stagnation may follow. In Model 1-C, investment does grow but at a rate too slow to absorb that part of growing capacity output that is not absorbed by consumption. This will also cause excess capacity to pile up year by year, which means that even the indicated lower-than-required growth rate probably cannot be maintained.

A Graphic Representation of the Growth Process[9]

With a few modifications, the familiar Keynesian short-run saving-investment figure for the determination of the equilibrium level of output can be converted into a figure that illustrates the growth process described above in tabular form. In Figure 21-1, the intersection of S and I_1 defines an original equilibrium level of output at Y_1.[10] The short-run analysis suggested that the equilibrium level of output would remain at Y_1 as long as neither the saving nor the investment function shifted. With no such shift, investment would be just equal to the leakage into saving at the given level of income and output; whatever part of income was not spent for consumption would be offset by an equal (planned) amount spent for investment. With no change in total spending, there would be no change in the equilibrium level of income and output. As we now know, this is an unrealistic conclusion except for very short-run situations, because it neglects the capacity effect of that part of spending composed of net investment. Thus, if we can incorporate this capacity effect into the figure, it can be used to illustrate the growth process in which capacity output does rise and in which full use of capacity is possible only if the amount of investment rises as required per time period.

Let us assume, as in Table 21-1, that $Y_r = Y_p$ at the output level Y_1 in year 1. Net investment for the year is I_1, which raises the capital stock at the beginning of year 2 by the same amount, so Y_p of year 2 exceeds Y_p of year 1. How large this increase in capacity output will be depends on I_1 and σ. The year's net investment is, of

[9]The figures in this section follow those developed by H. Pilvin in "A Geometric Analysis of Recent Growth Models," in *American Economic Review*, Sept. 1952, pp. 594–99.

[10]This portion of the figure differs from the familiar Keynesian figure only in that the short-run nonproportional saving function has been replaced by a long-run proportional function. As drawn, MPS = APS, in contrast to the short-run nonproportional function, in which MPS > APS.

course, represented by the height of the I_1 function above the horizontal axis. The value of σ is represented by the reciprocal of the slope of the E_1 function. The higher the value of σ, the less will be the slope of the function. If $\sigma = 3/4$, the slope will be $4/3$; if $\sigma = 1/4$, the slope will be 4, and in the special case in which $\sigma = 1$, the slope will be 1 (a 45° line), with the increase in capacity output just equal to net investment. With the slope of E_1 in Figure 21-1, I_1 raises productive capacity by Y_1Y_2. This is determined graphically by moving along the I_1 curve from S_1 to the intersection with the E_1 curve and dropping a vertical line to locate Y_2 on the output axis.

For year 1 we assumed that $Y_r = Y_p$, but now Y_p of year 2 exceeds Y_p of year 1. If there is to be full utilization of the increased productive capacity in year 2, Y_r must rise by an amount equal to the rise in Y_p. This in turn requires that aggregate spending rise by the same amount. With a stable saving function, the required rise in aggregate spending can occur only if investment rises in year 2 by the amount necessary to offset the rise in saving at the higher level of output. In other words, if Y_r is to rise from Y_1 to Y_2, net investment must rise from I_1 to I_2 to offset the rise in saving from S_1 to S_2 that will accompany a rise in output to Y_2. If such a rise in investment occurs, the economy will then operate at full capacity in year 2; $Y_r = Y_p$ at the output level Y_2. The process then repeats itself. Investment of I_2 during year 2 raises the capital stock at the beginning of year 3 by the amount I_2. With an unchanged value for σ, this increase in capital stock raises the productive capacity of the economy in year 3 from Y_2 to Y_3, as indicated by the E_2 line. Again, to make full use of the expanded productive capacity, investment will have to rise from I_2 to I_3 in year 3.

As was brought out by Table 21-1, equilibrium growth requires that the absolute increase in investment in each period exceed the absolute increase in investment of the preceding period. In terms of Figure 21-1, $I_3I_2 > I_2I_1$ and $I_4I_3 > I_3I_2$. With a constant value for σ, the increase in capacity output in each period also exceeds that of the preceding period, or $Y_3Y_2 > Y_2Y_1$ and $Y_4Y_3 > Y_3Y_2$. If numerical values were inserted on the axes of Figure 21-1, the percentage changes from Y_1 to Y_2 and Y_2 to Y_3 and from I_1 to I_2 and I_2 to I_3 would prove to be equal to $s\sigma$, the slope of the S function multiplied by the reciprocal of the slope of the E function.

The Addition of a Theory of Investment

The model, as developed so far, shows the rate at which investment *must* grow if aggregate spending is to grow at the rate needed to provide full utilization of a growing capital stock. Thus, column (7) of Table 21-1 showed the dollar amounts of investment, and Figure 21-1 showed the successive positions of the autonomous investment curve that must be attained if there is to be full utilization of productive capacity period after period. However, indicating what *must* happen in order to meet this objective tells us nothing about what *will* happen. There is nothing within the model itself to indicate what the actual value of investment will be period after period. It is specifically a model of this kind that Domar developed. He chose an approach that did not enter into the question of what determines the rate of investment but limited itself to identifying what the rate of investment would have to be period after period in order to provide an equilibrium rate of growth.

We may now, if we like, expand the Domar-type model by adding a theory of investment such as the accelerator theory so that the model will then contain a theory of what investment will be in each period. In combination with the theory of consumption already in the model, this additon will give us a model that contains a theory to cover both components of aggregate spending. It is specifically the inclusion of an accelerator theory of investment that, in a formal

sense, is the major difference between the approach taken by Harrod and that taken by Domar. Because these two approaches are essentially the same apart from this, they are commonly lumped together under the heading of the Harrod-Domar model. Still the difference is an important one whose full implications can be seen only by developing Harrod's approach in detail. We will not do that here, but we can gain some understanding of the effect of the addition of the accelerator investment equation to the model developed so far merely by comparing Figure 21-1, which showed what investment must be if the system is to follow an equilibrium growth path, with Figure 21-2, which includes a theory of what investment will be.

One form of the accelerator theory of investment makes investment of any period equal to the accelerator times the change in the level of output between the current and the preceding period, or, as an equation, $I_t = w(Y_t - Y_{t-1})$.[11] This is the version presented in Chapter 11, the one used by Harrod, and the one on which Figure 21-2 is based. In this figure the change in output from period 1 to period 2 is given by the distance between Y_1 and Y_2 on the income or output axis. How much investment this induces in period 2 depends on the size of the

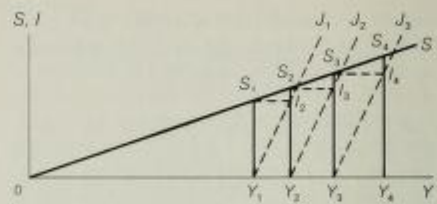


FIGURE 21-2
The growth process with an accelerator theory of investment

accelerator, and this is shown diagrammatically by the slope of the line labeled J_1. This and the other J lines are all drawn with a slope of 2 to conform with the assumption that the accelerator is 2. Thus, given a rise in output from Y_1 in period 1 to Y_2 in period 2, we find induced investment in period 2 to be equal to the vertical distance I_2Y_2, or equal to twice the change in output, Y_1Y_2. The figure is so drawn that the rise in output, Y_1Y_2, is just sufficient to induce investment of I_2Y_2, which is just equal to the saving made available in period 2. A lagged saving function is assumed in which each period's saving equals the propensity to save times the preceding period's income. Thus, saving made available in period 2 is equal to S_1Y_1, which is matched in period 2 by investment of I_2Y_2.

The rate of growth of income and output from period 1 to period 2 is equal to $(Y_2 - Y_1)/Y_1$. In Figure 21-2 the same rate of growth of income is maintained in succeeding periods, so that $(Y_2 - Y_1)/Y_1 = (Y_3 - Y_2)/Y_2$ and so forth. If income grows from period to period in the way described by the figure, the economy may be said to be moving along an equilibrium path. The only growth rate that will carry the system along such a path is the particular rate that is consistent with the values of the accelerator and propensity to save adopted in drawing the figure. Harrod calls this particular growth rate the *warranted* rate of growth, the term "warranted"

[11] It may be noted that algebraically the accelerator w, or K/Y, is the reciprocal of σ, or Y/K, the output–capital ratio or average productivity of capital, which appears above in the analysis that follows Domar's approach. If we rewrite $I_t = w(Y_t - Y_{t-1})$ in the simpler form, $I = w(\Delta Y)$, it will be seen that Domar's $\Delta Y = \sigma I$ may in turn be rewritten as $I = (1/\sigma)\Delta Y$—that is, in a form algebraically equal to Harrod's acceleration investment equation. However, although w and $1/\sigma$ are equal algebraically, their economic meanings are quite different. The rewritten Domar equation cannot be interpreted like the Harrod equation, in which I is made a function of ΔY and w refers to the ratio that business persons desire to establish between I and ΔY. No such intended or desired relationship between I and ΔY is present in the Domar approach. The equation from that model as originally given, $\Delta Y = \sigma I$, expresses the relationship that Domar employs—namely, a simple production function in which the change in the period's capacity output is technologically equal to some fraction of the period's net investment.

suggesting that the given combination of accelerator and propensity to save "warrant" the rate of growth indicated. If the propensity to save were higher (a greater slope to the S curve), it could be seen that a higher rate of output growth would be required to keep the economy moving along an equilibrium growth path. A smaller accelerator (a lesser slope to the J curve) would have a similar effect. In the same way, a smaller propensity to save or a larger accelerator would each mean that a lower rate of output growth would be required to keep the economy moving along an equilibrium growth path.

This equilibrium growth path has been commonly described as a "razor's edge," for the slightest departure of the actual growth rate from the equilibrium, or warranted, rate throws the economy off the equilibrium growth path into either "stagnation" or "exhilaration." To illustrate, let us assume that income has risen by less than Y_1Y_2 from period 1 to period 2. Induced investment in period 2 would then be less than I_2Y_2 and thus less than the amount of saving forthcoming in period 2. This means a deficiency of spending, which means excess capacity, or a situation in which the existing capital stock is greater than that desired by business at the then current level of output. Paradoxically, this excess capacity results from the fact that business did not invest enough or acquire enough additional capacity. An appropriately higher level of investment would have prevented the deficiency of demand and the appearance of excess capacity. However, the fact is that excess capacity has appeared, and this causes a contraction of income and output, which in turn, via the accelerator, causes a further reduction in investment. But then saving again exceeds investment, and there is a still greater deficiency of aggregate spending. Thus, once given an initial divergence in which the actual growth rate falls below the equilibrium, or warranted, growth rate, the forces at work push the system, not back toward the equilibrium growth rate but farther and farther away from it. In the event of such a divergence, the result is deep, prolonged depression or stagnation.

In the case of the opposite divergence, let us suppose that output has risen by more than Y_1Y_2 from period 1 to period 2. Induced investment in period 2 would then be greater than I_2Y_2 and thus more than the amount of saving made available in period 2. This means an excess of spending, which in the present context means a capital shortage, or a situation in which the existing capital stock is less than that desired by business at the then current level of output. As in the case above, there is a paradox here: The capital shortage results from the fact that business invested too much. An appropriately lower level of investment would have prevented the excess spending that resulted in the capital shortage. However, once a capital shortage appears, investment rises as business tries to make up for the shortage. Investment again exceeds saving, and there is a still greater excess of spending. Once given the divergence in which the actual rate exceeds the equilibrium, or warranted, rate of growth, forces push the system farther and farther away from the equilibrium growth rate. Where the divergence is an actual rate above the equilibrium rate, the result is secular exhilaration.

Because the actual rate of growth is subject to all sorts of influences, it can be expected that it will depart from the equilibrium rate of growth that is dictated by the values of the propensity to save and the accelerator. The picture suggested by Harrod's formulation is thus one of an economy with an inherently unstable pattern of growth, marked by a tendency toward long periods of boom or stagnation. For, as we have seen, once the economy departs from the equilibrium growth path, it moves farther and farther away over time. It is a focus on this alleged tendency toward secular instability that is the most striking feature of Harrod's formulation of the basic capital-accumulation theory known as the Harrod-Domar theory.

The Employment of Labor

The equilibrium, or warranted, rate of growth in the basic Harrod-Domar theory is the rate of growth that provides full utilization of a growing capital stock or the rate that results in neither shortage nor excess of capital. Although capital is the only factor of production that is explicitly considered, the model, of course, recognizes that labor too is needed in order to produce. Labor as a factor of production is, however, brought in through the assumption of fixed factor proportions, so it may be said that the amount of labor employed is indicated by the amount of capital in use. In other words, labor and capital are assumed to be perfectly *complementary* in the production process, but the focus of the model is on the stock of capital. The model also assumes constant returns to scale, so that, for example, a 1-percent increase in the amount of labor and capital will result in a 1-percent increase in the economy's capacity output.

If the economy grows at the equilibrium or warranted rate, the amount of labor employed will grow at the same rate, in view of the assumption of perfect complementarity. This suggests that there will be full employment of labor only if the warranted growth rate is equal to the growth rate of the labor force and only if the actual growth rate corresponds with the warranted growth rate. In the previous section we noted that the actual and warranted rates may be expected to diverge, and here it is to be noted that the warranted growth rate and the labor force growth rate may also be expected to diverge, given the essentially demographic factors that determine the long-run growth of the labor force.

If the actual growth rate equals the warranted growth rate but the warranted growth rate exceeds the labor force growth rate, the result is accumulation of idle plant and equipment, for there would be more new capital added than new workers to operate it. Since business persons quite plainly will not long continue to produce new plant and equipment that cannot be utilized for want of workers, the ceiling to the actual rate of growth in this case would appear to be a rate set by the growth of the labor force. The warranted rate might be 4 percent when the labor force growth rate is only 1 percent. Given the model's assumption of fixed proportions between labor and capital, the ceiling to the actual rate of growth of output would then be 1 percent.

The actual growth rate would not be limited to this 1-percent ceiling if we added to our illustration an allowance for the influence of technological advance. Technological advance of a labor-saving nature enables a smaller amount of labor and an unchanged amount of capital to produce an unchanged amount of output. In a growth context, this means that capital may now grow at a faster rate than the labor force because technological advance is continuously releasing some labor from the existing capital stock, which *in effect* is equivalent to a rise in the rate of growth of the labor force. The process may be viewed as one in which the labor so released becomes available for employment with new capital goods that would not have been produced if this labor had not been made available by labor-saving technological advance. If this rate of technological advance is, say, 2 percent and the rate of growth of the labor force is 1 percent, the ceiling growth rate becomes 3 percent instead of the 1-percent rate set by the actual growth of the labor force alone. However, this ceiling is still below the 4-percent rate assumed as the warranted rate, so the actual rate must still fall below the warranted rate, with the consequences of economic stagnation noted earlier.

In the opposite case, the labor force growth rate may be greater than the warranted growth rate. For example, if the propensity to save were relatively low and the investment of the amount of income saved permitted only a 2-percent growth rate for the capital stock, a labor force growth rate of more than 2 percent would mean

insufficient plant and equipment to employ the full labor force. The ceiling to the actual rate of growth would then be imposed by the warranted rate. With the labor force growth rate greater than this rate, the rate of growth of the capital stock would not be sufficient to provide full employment for the growing labor force. The dual objective of full utilization of both labor and capital cannot be satisfied under the conditions described, for labor and capital are employed in fixed proportions, and the labor force in this situation is simply growing faster than the capital stock. In short, full employment of labor becomes impossible.

These and other conclusions reached in the preceding discussion of the Harrod-Domar theory follow at least partially from the fact that that theory assumes fixed factor proportions. If we were to drop the assumption that labor and capital are combined only in this way, it would no longer necessarily follow in the situation just described that full employment of the labor force would become impossible. If the labor force growth rate exceeds the warranted growth rate, the ratio of labor to capital in the production process might increase so as to permit full employment of the labor force. From a much broader point of view, if we drop the assumption of fixed factor proportions, it is no longer meaningful to say that we can explain the economy's growth rate in terms of the propensity to save and the productivity of capital. With the possibility of variable proportions, the approach to an explanation of the growth rate must be one that brings the labor force explicitly into the analysis.

Once we introduce this possibility of variable factor proportions into the picture, we leave the world of Harrod-Domar. As noted in the introduction to this chapter, it is the adoption of the assumption of variable factor proportions that characterizes a body of more recent growth theory. The economists who returned to this neoclassical assumption of substitutability between factors built a new type of growth theory that differs basically from the Harrod-Domar type; it is generally known as neoclassical growth theory.

NEOCLASSICAL GROWTH THEORY[12]

In one sense, neoclassical growth theory stands at an opposite extreme from Harrod-Domar. In place of the Harrod-Domar assumption of a single production process that imposes a fixed ratio between capital and labor is the assumption of an indefinitely large number of production processes, one shading off from another in a way that permits any combination of labor and capital to be employed. Capital is thus regarded as a unique, abstract agent of production that can be adjusted at any time to absorb into employment a labor force of any size. With the combination of labor and capital capable of varying in this way, it follows that, instead of the fixed ratio between output and capital employed

[12] A sizeable literature has developed since the mid-fifties. Among the major contributions are R. M. Solow, "A Contribution to the Theory of Economic Growth," in *Quarterly Journal of Economics*, Feb. 1956, pp. 65–94; "Technical Change and the Aggregate Production Function," in *Review of Economics and Statistics*, Aug. 1957, pp. 312–20; T. W. Swan, "Economic Growth and Capital Accumulation," in *Economic Record*, Nov. 1956, pp. 334–61, and E. S. Phelps, "The New View of Investment: A Neoclassical Analysis," in *Quarterly Journal of Economics*, Nov. 1962, pp. 548–67. For less technical and nonmathematical expositions of neoclassical growth theory, see H. G. Johnson, "The Neo-Classical One-Sector Growth Model: A Geometrical Exposition and Extension to a Monetary Economy," in *Economica*, Aug. 1966, pp. 265–87, and J. E. Meade, *A Neo-Classical Theory of Economic Growth*, Oxford Univ. Press, 1961. See also D. Hamberg, *Models of Economic Growth*, Ch. 2, and H. Y. Wan, Jr., *Economic Growth*, Ch. 2.

by Harrod and Domar, the output–capital ratio is also capable of varying continuously. Thus, the larger the labor force absorbed into employment with a given stock of capital, the greater will be the output–capital ratio, or the productivity of capital, and the smaller will be the output–labor ratio, or the productivity of labor. Similarly, the smaller the labor force absorbed into employment with a given stock of capital, the lower the productivity of capital and the higher the productivity of labor. These results follow simply as a matter of diminishing returns.

The neoclassical theory also differs from Harrod-Domar in its assumption of automatic full utilization of capital and labor. The question of whether or not the amount of saving generated by an economy with factors fully utilized will be matched by an equal amount of planned investment, a requisite to continued full utilization of factors, is answered in terms of classical theory: All economic activity is carried out in conditions of perfect competition, with flexible prices of inputs and outputs serving to balance supply and demand in all markets along the lines discussed in Chapter 14. Output then depends simply on the supply of inputs, for all inputs available will find employment. With the question of full utilization of inputs resolved in this way, the neoclassical theory's focus is on the growth path that will be followed by a system whose labor and capital resources remain fully utilized as the quantity of these resources grows over time. For this reason, there is no need to distinguish in this discussion between the growth rate of the economy's capacity, or potential, output and the growth rate of its actual, or realized, output, for the latter becomes the same as the former.

The Rate of Output Growth Without Technological Progress

The rate at which the output of the economy grows depends basically on the rate at which its capital stock, labor force, and technological know-how grow over time. The relationship for any particular period of time may be simply expressed in the form of the following production function:

$$Y = f(K, L, A)$$

in which K is the capital stock, L the labor force, and A an index of technological know-how whose magnitude will grow at some rate over time. As a first step, however, we will temporarily simplify the relationship by assuming that no technological progress takes place. The production function is then

$$Y = f(K, L)$$

In this simplest case, how does Y vary as K and L vary, other things being equal? In the Harrod-Domar theory, if both K and L increased by 1 percent, Y would increase by 1 percent. The neoclassical theory indicates the same result, for both theories assume a production function with constant returns to scale. However, in Harrod-Domar, a 1-percent increase in L with no increase in K (and with all the existing K already in full use) would mean no increase in Y because of the fixed proportion between L and K, whereas in neoclassical growth theory the same set of conditions would lead to some increase in Y because the enlarged L could be absorbed into employment with an unchanged K.

If the increase in L is not very great in the period in question, the increase in Y will be approximately equal to the increase in L times the marginal physical product of L, or $\Delta Y = \text{MPP}_L \cdot \Delta L$ in which MPP_L is the marginal physical product of labor, or the increase in Y that accompanies a unit increase in L with K held constant. If we had assumed an increase in K with no change in L, under the same assumptions we would have had $\Delta Y = \text{MPP}_K \cdot \Delta K$ in which MPP_K is the marginal physical product of capital, or the increase in Y that accompanies a unit increase in K with L held constant. Finally, for changes in both K and L in a given time period, we may write

$$\Delta Y = \mathrm{MPP}_K \cdot \Delta K + \mathrm{MPP}_L \cdot \Delta L$$

Dividing both sides by Y, we have

$$\frac{\Delta Y}{Y} = \left(\frac{\mathrm{MPP}_K}{Y}\right) \Delta K + \left(\frac{\mathrm{MPP}_L}{Y}\right) \Delta L$$

which may also be written as

$$\frac{\Delta Y}{Y} = \left(\frac{\mathrm{MPP}_K \cdot K}{Y}\right) \frac{\Delta K}{K} + \left(\frac{\mathrm{MPP}_L \cdot L}{Y}\right) \frac{\Delta L}{L} \qquad [1]$$

If we recall the assumption noted earlier of perfect competition in all markets and now adopt the marginal productivity theory of factor pricing, each unit of a factor will be paid its marginal product, and the total earnings of capital and labor will be equal to $\mathrm{MPP}_K \cdot K$ and $\mathrm{MPP}_L \cdot L$, respectively. Given that factors are paid their marginal products, the total earnings of capital and labor will exactly absorb the total output in the case of the present production function with constant returns to scale. That is,

$$\mathrm{MPP}_K \cdot K + \mathrm{MPP}_L \cdot L = Y$$

Since

$$\frac{\mathrm{MPP}_K \cdot K}{Y} + \frac{\mathrm{MPP}_L \cdot L}{Y} = \frac{Y}{Y} = 1$$

we may substitute b for the first term on the left and $(1 - b)$ for the second term and rewrite Equation 1 in the following form:

$$\frac{\Delta Y}{Y} = b\left(\frac{\Delta K}{K}\right) + (1 - b)\frac{\Delta L}{L} \qquad [2]$$

The magnitude of b indicates the proportion of the total product or of total income that would be received as a return on capital if capital were paid its marginal product. This is the same as saying that b measures the elasticity of output with respect to changes in the amount of capital used.[13] If $b = 0.25$, we can say either that the owners of capital would receive 25 percent of the economy's income if capital earned a return equal to its marginal product or that a 1.0-percent increase in the amount of capital in use would produce a 0.25-percent increase in output. These amount to two ways of saying the same thing. The same kind of statement may, of course, be made for labor by making the appropriate substitution in the second preceding sentence.[14]

Assuming a value for b, say 0.25, we may read from Equation [2] the percentage change in output that will follow from a given percentage change in capital, labor, or both. If both K and L rise by 3 percent, output also rises by 3 percent, for the underlying production function is one with constant returns to scale. In this case, we have $\Delta Y/Y = 0.25 \times 3 + 0.75 \times 3 = 3$. However, note that capital and labor do not contribute equally to the growth in output, even though each factor grows at the same rate. In the illustration, 2.25 percentage points of the 3-percent growth rate are due to labor, and the remaining 0.75 percentage point is due to capital. Labor is given three times the weight of capital, for the weights correspond with their output elasticities, and these show that a 1-percent increase in labor will produce a percentage increase in output three times as large as a 1-percent increase in capital. Thus, a 3-percent increase in K and a 1-percent increase in L would indicate a 1.5-percent increase in output, but a 1-percent increase in K and a 3-percent increase in L would indicate a 2.5-percent increase in output.

The results derived here are very different from those indicated by the Harrod-Domar theory. In terms of the numerical example above, if we started from a position of full utilization of labor and capital, a 3-percent growth rate for K and a 1-percent growth rate for L, or vice versa, would in both cases mean a ceiling growth rate for Y of 1 percent. This again is

[13] The general concept of elasticity refers to a percentage change in one variable divided by the percentage change in another. If L is not changing, $b = \Delta Y/Y \div \Delta K/K$, or the percentage change in Y over the percentage change in K, or b is an elasticity.

[14] That $(1 - b)$, or $(\mathrm{MMP}_L \cdot L)/Y$, is equal to labor's share of total output follows the analysis of Chapter 13. We saw there that under competitive conditions employers will hire labor up to the point at which $\mathrm{MPP}_L \cdot P = W$, or $\mathrm{MPP}_L = W/P$. Substituting W/P for MPP_L in $\mathrm{MPP}_L \cdot L/Y$ gives us WL/PY, or the total money wage income of labor as a fraction of the total money income for the economy.

the necessary result if we rule out the substitutability between factors that is not ruled out in neoclassical growth theory. In a more general way, we may bring out the basic difference between the Harrod-Domar and neoclassical theories by rewriting Equation [2] in a way that incorporates the Harrod-Domar theory. Since $\Delta K = I = sY$ and since $Y/K = \sigma$ in which σ is the symbol used earlier for the output–capital ratio, or the productivity of capital, we may rewrite Equation [2] to read

$$\frac{\Delta Y}{Y} = b(s\sigma) + (1 - b)\frac{\Delta L}{L}$$

Recall that $s\sigma$ defined the equilibrium rate of growth in the Harrod-Domar theory, or the growth equation was simply $\Delta Y/Y = s\sigma$. In the neoclassical theory, with its allowance for substitutability between factors, the rate of growth of labor has its own influence on the rate of growth of output. The propensity to save, s, and the productivity of capital, σ, whose product equals the rate of growth of capital, are in themselves no longer sufficient to explain the rate of growth of output. Actually, the rate of growth of capital not only is insufficient in itself to explain the rate of growth of output but may be of far less importance than the labor force in this regard. This follows from the fact that the influence of capital accumulation on the growth rate depends heavily on the size of b, and b may be relatively small. To trace through a numerical illustration, suppose that the economy's output is growing at 3 percent over a period when s is 0.10 and σ is 0.30, so that $s\sigma$ equals 0.03, or 3 percent. If the economy became twice as thrifty and raised s to 0.20, would the rate of growth of output also double, rising to 6 percent? According to the Harrod-Domar theory, a doubling would result, but here we find that it results only if b equals 1.[15] But b is actually in the neighborhood of 0.25 to 0.33, the range of the share of capital in total income. If b were 0.25, the doubling of the saving ratio would raise the growth rate of output only from 3.0 to 3.75 percent; if b were 0.33, such a doubling would raise the growth of output only from 3.0 to 4.0 percent. The fixed link between the rate of growth of capital and the rate of growth of output given by Harrod and Domar is replaced here by a link in which the rate of growth of output depends on the rates of growth of both capital and the labor force and, further, on the output elasticities of capital and labor.

[15] If b were to equal 1, the neoclassical equation would reduce to $\Delta Y/Y = s\sigma$, or to the Harrod-Domar equation. However, for b to equal 1 requires that there be a surplus of labor that is not at all substitutable for capital in production, so output can grow only at the rate of growth of capital, or $\Delta Y/Y = \Delta K/K = s\sigma$.

The Rate of Output Growth per Capita

Although there is obvious interest in how the rate of growth of aggregate output (or $\Delta Y/Y$) will vary with different rates of growth of labor and capital, of much greater interest is how the rate of growth of output per worker will vary under the same circumstances. If we subtract $\Delta L/L$ from $\Delta Y/Y$, we have the difference between the growth rate of aggregate output and the growth rate of the labor force, a difference that is approximately equal to the growth rate of output per worker. Therefore, we can convert Equation [2] into one that shows the growth rate of output per worker by subtracting $\Delta L/L$ from both sides. The same subtraction will show the growth rate of output *per capita*, if the rate of growth of population is the same as the rate of growth of the labor force. In what follows, if we assume these rates are equal, growth rates per worker may also be read as growth rates per capita. Making the indicated subtraction gives us

$$\frac{\Delta Y}{Y} - \frac{\Delta L}{L} = b\left(\frac{\Delta K}{K} - \frac{\Delta L}{L}\right) \qquad [3]$$

The growth rate of output per worker ($\Delta Y/Y - \Delta L/L$) is equal to the growth rate of capital per worker ($\Delta K/K - \Delta L/L$) weighted by the elasticity of output with respect to capital. In order for the growth rate of output per worker to

be above zero, the equation shows that there must be an increase in the stock of capital per worker—that is, a growth rate of capital greater than the growth rate of the labor force. For example, suppose that $\Delta Y/Y$ equals 2 percent and $\Delta L/L$ equals 1 percent, a combination that indicates a 1-percent rise in output per worker per period. The right-hand side of Equation [3] shows that, to secure this 1-percent rise in output per worker per period, the rate of growth of capital must be 5 percent if we assume a value for b of 0.25, or 4 percent if we assume a value for b of 0.33. Thus, given a value of b in this range, we find in the present case that a rate of growth of capital four to five times the rate of growth of the labor force is required to produce a 1-percent rate of growth in output per worker per period.[16]

Is it possible for the capital stock to grow at a higher rate than the rate of growth of the labor force period after period, which, according to Equation [3], is a prerequisite to a rising output per worker period after period? A rate of growth of the capital stock greater than the rate of growth of the labor force must result in a decline in the average productivity of capital, or in the output–capital ratio—diminishing returns makes this so. Correspondingly, there must result a rise in the average productivity of labor, or in the output–labor ratio. The decline in the average productivity of capital amounts to a fall in the rate of return on capital. This may be seen through a variation of Equation [3]. The rate of

If b is 0.25, growth rates of 5 percent for capital and 1 percent for labor produce a 3-percent *decline* in the productivity of capital per time period. Whatever the level from which the decline starts, if our periods are years, this means that in less than fifteen years the return on capital will have fallen to less than two-thirds of its previous level. As a matter of fact, Equation [4] indicates that for any value of b less than 1 and for any values of $\Delta K/K$ and $\Delta L/L$ in which $\Delta K/K > \Delta L/L$, the rate of growth of productivity of capital will be negative and the rate of return on capital will be a declining figure. Under these conditions, capital will not grow more rapidly than labor. The very-long-run adjustment in the present case will be one in which both capital and labor ultimately grow at the same rate.

However, we saw that growth in output per worker cannot be achieved unless there is growth in capital per worker, and here we see that the very-long-run tendency is for capital to grow at the same rate as labor. Since we know from the historical record both that capital has grown faster than labor and that there has been growth in output per worker, there appears to be something wrong with the formulation above. It is actually correct, given its assumptions, but one of the critical assumptions it makes is that no technological progress occurs. If this were indeed the case, the neoclassical model would lead to a so-called steady state

of output, but there are difficult questions in establishing precisely how it works out its effect on the output growth rate. We will touch on this issue in the next section. As a first step, however, it is helpful to simplify things by viewing technological progress as a force that merely raises the growth rate of output above what it otherwise would be. It is thus viewed as a factor that is independent of the labor and capital factors with which it actually collaborates to some degree in production. With this simplification, we may refer to the rate of technological progress as a 2-percent rate if, in the course of the time period, such progress makes possible the production of an aggregate output 2 percent greater than would otherwise have been possible with an unchanged amount of labor and capital. Designating the rate of technological progress by $\Delta A/A$, we have in place of Equation [2] the following:

$$\frac{\Delta Y}{Y} = \frac{\Delta A}{A} + b\left(\frac{\Delta K}{K}\right) + (1-b)\frac{\Delta L}{L} \qquad [5]$$

and in place of Equation [3] the following:

$$\frac{\Delta Y}{Y} - \frac{\Delta L}{L} = \frac{\Delta A}{A} + b\left(\frac{\Delta K}{K} - \frac{\Delta L}{L}\right) \qquad [6]$$

It is no longer true that growth of output per worker can occur only if there is growth in capital per worker. We now find that even if $\Delta K/K = \Delta L/L$, $\Delta Y/Y - \Delta L/L$ will be greater than zero by an amount equal to $\Delta A/A$. Indeed, the growth rate of output per worker will be still higher if $\Delta K/K > \Delta L/L$, but this is not a prerequisite to a positive growth rate in output per worker as it was before technological advance was brought into the equation.[17]

[17] In the same way that Equation [6] replaces Equation [3], we may replace Equation [4] with the following:

$$\frac{\Delta Y}{Y} - \frac{\Delta K}{K} = \frac{\Delta A}{A} + (1-b)\left(\frac{\Delta L}{L} - \frac{\Delta K}{K}\right)$$

which indicates that the productivity of capital need not necessarily fall in the event the capital stock grows at a faster rate than the labor force. The innovations and technical changes that give $\Delta A/A$ its value may be sufficient to raise the productivity, or real return on new capital, more rapidly than it is reduced by a growing ratio of capital to labor—that is, by diminishing returns.

The rate of growth of output per worker thus depends on the capital stock per worker and the rate of technological advance. These two sources are all-inclusive if we view technological advance as a residual, or catchall, that includes all sources of growth other than the rise in capital per worker. Since the growth rate of output per worker is $\Delta Y/Y - \Delta L/L$ and since we know from Equation [3] that the growth rate of output per worker due to the growth in capital per worker is $b(\Delta K/K - \Delta L/L)$, the ratio of the latter to the former gives the proportion of the total growth rate of output per worker that is due to the growth in capital per worker. The proportion of the total growth rate of output per worker that is due to technological advance is then obtained as a residual, or the difference between this proportion and 1. The proportion due to growth in capital per worker, here designated by e, may be expressed as follows:

$$e = \frac{b\left(\frac{\Delta K}{K} - \frac{\Delta L}{L}\right)}{\frac{\Delta Y}{Y} - \frac{\Delta L}{L}}$$

If $\Delta K/K = \Delta Y/Y$, then $e = b$. We have seen that a realistic value for b is in the range from 0.25 to 0.33, so in this case e would be 0.25 to 0.33. To the extent that $\Delta K/K$ is smaller than $\Delta Y/Y$—and this appears to have been the case for the last sixty years—the proportion e declines to less than 0.25 for b of 0.25 and to less than 0.33 for b of 0.33. Figures such as these have led economists working in this area to conclude that *less than one-third of the growth rate of output per worker over the years from the turn of the century can be attributed to the rise in capital per worker. Over two-thirds of the growth rate of output per worker has therefore to be attributed to all the other factors covered by the catchall called technological advance.* This same conclusion may be expressed alternatively by stating that *less than one-third of the growth rate of total output can be attributed to the growth rate of the labor and capital inputs, so over two-thirds of the growth rate of total out-*

TABLE 21-3
Selected growth rates, 1929–60 and subperiods (average annual rates)

PERIOD	GROSS CAPITAL STOCK $\Delta K/K$	POTENTIAL MAN-HOURS $\Delta L/L$	POTENTIAL GNP $\Delta Y/Y$	TECHNOLOGICAL ADVANCE $\Delta A/A$ WITH $b = 0.25$	TECHNOLOGICAL ADVANCE $\Delta A/A$ WITH $b = 0.50$
1929–60	2.0	0.7	3.1	2.1	1.7
1929–47	1.0	0.5	2.5	1.9	1.7
1947–60	3.6	0.8	4.0	2.5	1.8
1947–54	4.0	0.7	4.4	2.9	2.1
1954–60	3.1	0.8	3.5	2.1	1.6

SOURCE: R. R. Nelson, "Aggregate Production Functions and Medium-Range Growth Projections," in *American Economic Review*, Sept. 1964, Tables 1 and 2, pp. 577 and 579.

put has to be attributed to increased output per unit of labor and capital inputs, or again to all the factors included under the catchall heading of technological advance.[18]

These are rather startling conclusions, and it is worth looking briefly at some of the specific estimates that provide the basis for such conclusions. The first three columns of Table 21-3 give the historical record of growth rates of the capital and labor inputs and aggregate output for 1929–60 and for subperiods.[19] The figures in the last two columns are found as residuals by substituting the figures from the first three columns in the following equation (which is the same as Equation [5] with the terms rearranged):

$$\frac{\Delta A}{A} = \frac{\Delta Y}{Y} - b\frac{\Delta K}{K} - (1 - b)\frac{\Delta L}{L}$$

$$2.1 = 3.1 - 0.25(2.0) - (0.75)0.7$$

Thus, on the assumption that b equals 0.25, for 1929–60 we find as shown a value for $\Delta A/A$ of 2.1 percent, or we have the conclusion noted above: Less than one-third of the growth rate of total output is attributable to the growth of the labor and capital inputs. The fraction of the total growth rate explained by the growth of labor and capital inputs is, of course, higher if b equals 0.50. However, even then this fraction remains below one-half for the 1929–60 period and for all the indicated subperiods.

If we separate out the fraction of the growth rate of output that is attributable to the growth of the capital stock alone, we are down to less than one-sixth for the 1929–60 period, assuming that b equals 0.25. The output growth rate was 3.1 percent, of which 0.5 percentage point (0.25×2.0) is explained by the growth of the capital stock. If we look at the two major subperiods into which 1929–60 has been divided,

[18]These estimates differ from those by Denison presented in Table 20-3. The lower figure of 47 percent he gives for output per unit of input follows from differences in classification, the major one being his inclusion in the labor input of an allowance for the improvement in the quality of the labor force. This is similar to what we will refer to in the next section as the "embodiment of technological advance."

[19]GNP, the measure of output, and man-hours, the measure of labor input, are given here as *potential* rather than *actual* rates. To avoid the distortions that would result from different degrees of slack in the economy at the various terminal dates, the actual GNP and man-hour figures are adjusted to what they would have been had the unemployment rate been 4.0 percent. The figures so adjusted give us what are designated as potential GNP and potential man-hours.

This is similar to the measure employed by Denison in deriving the estimates given in Table 20-3. Despite this, there are differences in the growth rates in the two tables, in part due to the difference in the aggregate in terms of which output is measured, GNP versus national income, and to the difference in the time periods covered.

the growth in the capital stock explains only one-tenth of the output growth rate for 1929–47 and about one-fourth of the output growth rate for 1947–60. To the earlier conclusion that labor and capital combined account for only about one-third of the growth rate of output, we may add the conclusion that capital accumulation by itself accounts for something like one-sixth of the growth rate of aggregate output during the 1929–60 period.

The Embodiment of Technological Progress

Technological progress was introduced into the preceding analysis in the simplest possible way—as a factor completely independent of the capital stock growth rate and the labor force growth rate. This way of bringing in technological progress amounts to viewing it as technical know-how falling like manna from heaven and somehow applying equally and impartially to all units of the economy's labor force and capital stock. Such technological progress, e.g., organizational improvements, is formally described as *disembodied*, for it yields its benefits without the necessity of embodiment in newly produced capital goods or newly trained or educated workers. As long as technological progress is disembodied, it is possible to assume in theory that the labor force is homogeneous, because the productivity of all workers, regardless of age, training, or education, will benefit proportionally from technological progress; and it is similarly possible to assume that all units of capital are homogeneous because all units of capital, regardless of age and design, will benefit proportionally from technological progress.

By contrast, *embodied* technological progress is a kind that must be physically incorporated in newly produced capital goods or newly trained or educated workers before it can contribute to the rate of growth of the economy's output. (For example, jet engines required embodiment in newly produced capital goods, namely, jet airplanes.) Capital can then no longer be assumed to be homogeneous. The capital stock becomes a mixed stock of different "vintages." Because they embody more technological progress, newer machines are more productive than older ones. Similarly, the labor force can no longer be assumed to be homogeneous. Like units of the capital stock, different individuals in the labor force are of different vintages, distinguished by age and training or education. Individuals of the current vintage are then more productive than individuals of earlier vintages.

Although embodied technological advances refer to those that are embodied in either the capital stock or the labor force—that is, in "physical capital" or "human capital"—we will in what follows limit attention to technological advance of a kind that either is or is not embodied in physical capital. Advances of this kind have been described as "design" or "organizational," the former requiring new capital goods and the latter requiring only new procedures or methods.

There is disagreement among economists working in this area as to what part of actual technological advance is design and what part is organizational. If it is true, as some argue, that the larger part is of a design nature and thus of a nature that must be embodied in new capital goods if its benefits are to be gained, then it also appears true that the real importance of the capital stock cannot be measured, as it was above, merely by the fraction of the growth rate of output that is directly explained by the growth rate of the physical stock of capital. We saw that for the period 1929–60 this fraction was a low one-sixth. Its real importance is much greater than this fraction indicates, for the capital stock is the vehicle through which technological progress is worked into the production process and without which the growth rate of output would be only a fraction of what it actually is. From this point of view, the capital

stock, which otherwise could be downgraded to a minor role as we saw above, reassumes the position of primacy that most people always thought it occupied.

A numerical illustration will bring out the difference between the earlier approach, in which technological advance was assumed to be completely disembodied, and the present approach, in which it is in part embodied. We assume an economy whose capital stock is equally divided into one-year through twenty-year vintages. All capital goods last exactly twenty years, so at the end of each year the capital goods then of twenty-year vintage come up for replacement. This amount will be 5 percent of the capital stock. We also assume that gross investment is equal to replacement investment, or net investment is zero. This means that each year's gross investment is equal to 5 percent of the capital stock, which in turn is a physical stock that does not change in amount from year to year.

Some part of the economy's technological advance is embodied in newly produced capital goods. Let us assume that technological advance is such that the capital goods produced each year show a 3.5-percent quality improvement, or productivity increase, over the capital goods produced during the preceding year. Any variable that grows at a 3.5-percent compound rate per annum will double in size in twenty years. Therefore, a capital good newly produced in year t and embodying the improvements of year t and all preceding years will have twice the productivity of the capital good of year $t - 20$ that it replaces. The comparison is, of course, between capital goods of equal real cost; the physical makeup of the old and the new may be completely different.

Since gross investment is by assumption equal to replacement investment and since replacement is understood to be the amount of capital goods that must be produced during the period in order to keep the total capital stock unchanged, we appear to have a case of a zero growth rate in the capital stock. However, even though each unit of capital is replaced at the end of twenty years by another unit of equal real cost, the replacement unit is twice as productive as the replaced unit and thus in effect is equivalent in real terms to two of the replaced units. For example, if the total capital stock is thought of as 100 units, this year's gross investment of 5 units for replacement equals the 5 units that wear out during this year and keeps the total capital stock unchanged at 100 units; but, since the 5 replacement units are equivalent in productivity to 10 of the replaced units, in effect the capital stock has been enlarged from 100 to 105 units. In our simple case, this increase would not be revealed to the statistician who measured merely the number of units or the physical quantity of capital without regard for the fact that the units produced each year, even if alike in appearance to units produced in preceding years, incorporate quality improvements not present in units produced in preceding years. An increase in the capital stock is nonetheless there but not in explicit form. In the present example, this implicit, or effective, increase of 5 in the capital stock, relative to the preexisting stock of 100, indicates what amounts to a 5-percent growth rate ($\Delta K/K = 5/100 = 5$ percent). Otherwise viewed, gross investment that is equal to replacement—that is, to 5 percent of the capital stock—is effectively gross investment equal to 10 percent of the capital stock, one half of which is 5 percent for replacement and the other half of which is, in effect, 5 percent for net investment.

On the basis of all the preceding assumptions and if we now add the final assumption that b equals 0.25, we find in this illustration that the part of technological advance that is embodied in the capital stock accounts for 0.25(5/100), or about 1.2 percentage points of the rate of output growth. This figure of 1.2 is derived from a whole series of assumptions, some quite arbitrary, and is likely to be far wide of the true figure, which is unknown. Nonetheless, let us

suppose for the moment that the true figure is at least one-half of the 1.2—that it is at least 0.6. What is relevant here is that whether we take 0.6 percentage point or a larger or even smaller figure, we still reach conclusions quite different from those reached in the previous section in which technological advance was completely disembodied.

There we presented some estimates for the 1929–60 period that showed that the output growth rate of 3.1 percent per year for this period was accounted for as follows: 0.5 by the growth rate of the labor force, 0.5 by the growth rate of the capital stock, and the residual of 2.1 by the catchall called technological advance. Some part of these 2.1 percentage points is undoubtedly made up of technological advance of a type that is embodied in the capital stock. On the assumption that the embodiment of technological advance in capital goods accounts for at least the 0.6 percentage point of the growth rate noted above, the 2.1-percentage-point residual is reduced to around a 1.5-percentage-point residual. This remains a catchall that includes disembodied technological advance, but it is now a conceptually narrower term that excludes the technological advance embodied in capital goods.

If 0.6 percentage point of the output growth rate could indeed be accounted for by the embodiment of technological advance in capital goods, the role of the capital stock in the growth process would be much more important than it appeared to be in the preceding section. Of the 3.1-percent growth rate of output for 1929–60, we saw that only 0.5 percentage point was accounted for by the growth in the quantity of capital goods. But suppose now that something like 0.6 percentage point of this growth rate is the result of embodied technological advance. This figure reflects the improvement in the quality of capital goods realized through modernization, and the improvement in the *quality* together with the growth in the *quantity* of capital goods may then account for over one-third of the growth rate of output. Again, the correct figure is not known, but if for the time period in question it was around 0.6 percentage point, it would mean that improvement in the quality of the capital stock was about as important as the growth in its quantity in explaining the rate of output growth.

As we have emphasized, this figure of 0.6 may understate or overstate the true importance of embodied technological advance and is submitted here for illustration only. Without knowing what the correct figure really is, many economists still feel, on the basis of available evidence, that the contribution of the capital stock to the rate of output growth comes in large part, if not primarily, from the role of capital as a vehicle for the embodiment of technological advance. There is, in sum, much more to the role of the capital stock in the growth process than the mere growth in the quantity of capital goods.

Although we chose to limit attention in this section to the question of the embodiment of technological advance in physical capital, it should be noted that the question of the embodiment of technological advance in "human capital" involves a parallel analysis. Just as the mere growth in the quantity of physical capital accounts for a part of the growth of output, so the mere growth in the quantity of human capital does the same. But, as in the case of the capital stock, over and above the mere growth in the amount of labor input is the question of the improvement in its quality as the result of the embodiment of education and training in people. Hard answers are not easy to come by, but it may well be that this is a more important factor in explaining the growth rate of output than is the embodiment of technological advance in the capital stock[20]

[20]Note that Denison's estimates for 1929–69 in Table 20-3 attribute 0.87 percentage point of the output growth rate to employment and hours and 0.41 to education. Education thus becomes about half as important as mere growth in labor input.

A CONCLUDING NOTE

In contrast to the Harrod-Domar theory, which assumes fixed proportions between labor and capital, the basic neoclassical theory treats the capital stock as a unique, homogeneous resource that can be reshaped and adjusted to be used with any quantity of labor. However, once embodied technological progress is brought into the neoclassical theory, as was done in the preceding section, the capital stock can no longer be viewed as homogeneous; each vintage of capital goods differs from every other vintage in terms of the technological advance that it embodies, the current vintage embodying the latest. The fact that we recognize that units of the capital stock differ according to vintage logically suggests that we should also recognize that the amount of labor required for units of capital of each vintage may differ. For example, capital of the oldest vintage now in use typically requires more labor per unit than capital of more recent vintage, since the older capital was not designed to hold down the amount of required labor to the degree that is true of more recent capital designed in times of relatively higher labor costs.

In addition to these differences in labor requirements between capital of different vintages, in practice the labor requirement per unit of existing capital of each vintage is not variable over a wide range, so the total of existing capital will similarly have a fairly specific labor requirement. Instead of the neoclassical assumption that the whole of the capital stock can be combined with the larger or smaller quantity of labor, we find that this assumption is strictly applicable only to capital goods that are in the design stage. At this stage, it is possible to choose, within the existing technology, between alternative forms that have different labor requirements. However, once the capital goods are constructed, the labor requirement is pretty much fixed for these goods, which constitute the latest vintage, as it was earlier for the capital goods that make up the earlier vintages.

For capital goods of vintages other than the current one, the Harrod-Domar assumption of fixed proportions between labor and capital may thus be not too unrealistic. However, that the proportion is not a strictly fixed one follows from the fact that it is subject to change in each year, for in each year there is the possibility of varying the labor–capital ratio via the labor requirement that designers have built into that year's capital goods. Although this cannot substantially change the average labor requirement per unit of the total capital stock in any one year, it can do so over a period of years, which is the time interval that is relevant to growth theory.

As is often the case with rival theories that in their simple forms present opposite extremes, a type of growth theory that combines features of both the Harrod-Domar and the neoclassical theories may be a better one than either of the two taken separately. Such theories have been developed in which the existing capital stock has a fixed labor requirement and the capital goods about to be built may be built with any of a variety of labor requirements. This type of theory sacrifices the uniformity and relative simplicity of the Harrod-Domar and neoclassical approaches, but, by providing a closer approximation to the complexity of the real world, it also helps improve our understanding of the process of growth in that extremely complicated real world.

chapter 22

Inflation: Definition and Effects

If subjects could be taken up in order of their importance, that of inflation would deserve a position in the first rather than the last part of a book on macroeconomics written during the seventies. However, a prerequisite to any study of the process and causes of inflation is an analytical framework within which such a study may be undertaken, and it was not until the development of the extended model in Part 4 of this book that such a framework became available to us.

Although the term "inflation" appeared a few times in earlier chapters, we paid no attention to it as a phenomenon in itself. In this and the following chapter we turn specifically to the phenomenon of inflation. Only a few of the large number of questions in the subject area of inflation can be considered here. In this chapter, we start off with a brief look at the definition of inflation. Everybody talks about inflation, but merely defining it is something of a problem. Inflation is rarely regarded as a blessing to society as a whole, although it can turn out to be that for some members of society. However, if it is a social evil, it is a greater or lesser evil depending on how harmful its effects are. The effects of inflation are covered in some detail in the balance of this chapter. Of critical importance, of course, is the question of what causes inflation, and it is to that major question that the following chapter will be primarily devoted.

INFLATION DEFINED

The obvious definition of inflation as a rising price level is far from unambiguous. One problem is that the presence or absence of inflation may depend in some periods on which of the available price indexes we use as a measuring stick of price change. For example, referring to the three major price indexes in Table 22-1, we find that for the period 1958–64 there was no inflation at all according to the wholesale price index, but there was what may be called a mild or "creeping" inflation according to the consumer price index and an inflation a little

stronger than this according to the GNP deflator. For the period since 1965, each of the three indexes shows persistently rising prices, but the year-to-year figures for the rate of price change are noticeably different in certain years, especially the percentage change in the wholesale price index relative to the percentage in the other two indexes.

In view of such differences, one cannot always say that there is or is not inflation or indicate the amount of it without reference to the price index on which one bases his conclusions. Because the GNP deflator has the broadest coverage and thus comes closest to the concept of the general price level of final goods and services, economists have come to

TABLE 22-1
The consumer price index, the wholesale price index, the GNP implicit price deflator and average annual rates of change in each for selected years, 1929–76

	CONSUMER PRICE INDEX 1967 = 100	AVERAGE ANNUAL RATE OF CHANGE	WHOLESALE PRICE INDEX 1967 = 100	AVERAGE ANNUAL RATE OF CHANGE	GNP IMPLICIT PRICE DEFLATOR 1972[a] = 100	AVERAGE ANNUAL RATE OF CHANGE
1929	51.3		49.1		32.9	
1933	38.8	−5.6	34.0	−6.9	25.1	−5.4
1940	42.0	1.1	40.5	2.5	29.1	2.1
1945	53.9	5.1	54.6	6.1	38.0	5.4
1948	72.1	10.2	82.8	14.9	53.1	11.7
1950	72.1	0.0	81.8	−0.5	53.6	0.4
1951	77.8	7.9	91.1	11.4	57.3	6.9
1956	81.4	0.8	90.7	−0.1	62.9	2.1
1957	84.3	3.6	93.3	2.9	65.0	3.3
1958	86.6	2.7	94.6	1.4	66.1	1.7
1964	92.9	1.2	94.7	0.0	72.7	1.5
1965	94.5	1.7	96.6	2.0	74.3	2.2
1966	97.2	2.6	99.8	3.3	76.8	3.3
1967	100.0	2.9	100.0	0.2	79.0	2.9
1968	104.2	4.2	102.5	2.5	82.6	4.6
1969	109.8	5.4	106.5	3.9	86.7	5.0
1970	116.3	5.9	110.4	3.7	91.4	5.4
1971	121.3	4.3	113.9	3.2	96.0	5.1
1972	125.3	3.3	119.1	4.6	100.0	4.1
1973	133.1	6.2	134.7	13.1	105.8	5.8
1974	147.7	11.0	160.1	18.9	116.0	9.6
1975	161.2	9.1	174.1	9.2	127.2	9.7
1976	170.5	5.8	182.9	5.1	133.9	5.3

SOURCE: Index numbers for 1929–74 from *Annual Report of the Council of Economic Advisers*, January 1977, and for 1975–76 from *Survey of Current Business*, U.S. Department of Commerce, July 1977. Rates of change computed by the author.

favor it as the best single guide to the measurement of inflation in the U.S. economy.[1]

Beyond this, whatever the price index chosen, there is a question of how rapidly that index must rise in order to qualify as inflation. For example, there is little argument that the rate of change of the GNP deflator over the years since the start of this decade qualifies each of these years as a year of inflation, but does the average annual rate of change of the GNP deflator from 1958 to 1964, a mere 1.5 percent, qualify this period as one of inflation? (The year to year rate varied from 0.9 to 2.2 percent over this period.) Some people regard inflation as something bad but consider a price level rising at a rate of around 1.5 percent as something good, believing that such a gradually rising price level assists in achieving and maintaining full employment and a satisfactory rate of economic growth. Such persons would apply the term "inflation" only to what they regard as an "unhealthy" or "excessive" rise in the price level. Other people agree that inflation is something bad, but they go further and label any rise in the price level as inflation, even the rise during the 1958–64 period, slow as it was.

Unless we adopt this latter viewpoint and describe any rise in the price level as inflation, we are forced to admit the possibility of a noninflationary rise in the price level. If we accept this, it then appears that there is some rise in the price level per period at which a price movement becomes inflationary. The problem is to determine how rapid the rise must be. Many writers tend to fall back on words such as "appreciable," "considerable," or "sizable," and some such question-begging is unavoidable. The only alternative is to resort to some fixed percentage rate, but the arbitrariness of any such figure would be worse than the ambiguous adjectives.

We encounter another definitional problem when we consider the time dimension. A 2-percent rise in the GNP deflator over any one quarter would, if sustained, amount to more than an 8-percent rise over a full year. An 8-percent rise in a year would be regarded by everyone as an appreciable and therefore inflationary rise in the price level. But suppose the 2-percent rise in one quarter were to reverse itself in the following quarter. Shall we describe the first quarter as a period of inflation? Or must an upward price movement be not only appreciable but prolonged in order to qualify as an inflationary price movement? If so, we must again fall back on question-begging qualifications such as "continuing," "persistent," or "sustained" rises in the price level, but again this is unavoidable. The only alternative is to resort to some fixed calendar period, but the arbitrariness of the selection of any such period would be worse than the adjectives.

In short, as long as one defines inflation in terms of price level changes, no fully satisfactory definition is available. However, this problem is not too serious, for what matters is not how fast and how long the price level must rise before the rise is called inflationary, but what causes the price level to rise in the first place and what are the consequences of different rates of price level change for the distribution of income, the level of output and

[1]However, because the GNP deflator is available only on a quarterly basis, it is not very suitable for policy purposes. Month-by-month observations on the movement of prices rely on the two other principal indexes in Table 22-1 or on others less familiar. For an introduction to the construction of price indexes and a description of the three major price indexes which are the ones in Table 22-1, see W. H. Wallace, *Measuring Price Changes: A Study of the Indexes*, 2nd ed., Federal Reserve Bank of Richmond, 1972. The Bureau of Labor Statistics has scheduled publication beginning in the fall of 1977 of a second consumer price index based on the cost of a market basket of goods and services for all urban households. The original index, that in Table 22-1 whose publication will continue, is based on the cost of a market basket for urban wage earners and clerical workers only. The new index extends coverage to about 80 percent of the total noninstitutional population from the approximately 45 percent of the total covered by the original index. For a description of the new index, see J. Shiskin, "Updating the Consumer Price Index—an overview," in *Monthly Labor Review*, July 1974, pp. 3–20, and *The Consumer Price Index: How Will the 1977 Revision Affect It?*, Report 449, 1975, Bureau of Labor Statistics, U.S. Department of Labor.

employment, the growth rate, and other critical variables used to measure the economy's performance. Recognizing the ambiguities our words contain, we will define inflation simply as *a persistent and appreciable rise in the general level of prices.*[2]

THE ECONOMIC EFFECTS OF INFLATION

Although some members of society gain from inflation, others get hurt, and the major measure of how badly they get hurt is the amount of their income and wealth that inflation takes away from them. In an economy in which the prices of everything, including the prices of assets and debt instruments, changed proportionally with the price level, nobody would be hurt by or benefit from changes in the price level. However, the real economy exhibits anything but such a condition, so inflation does provide gains to some and losses to others.

Of course, those who correctly anticipate inflation and find ways to protect themselves against the loss of income and wealth that would otherwise occur obviously do not get hurt in this way, but there are not too many able to do this. Actually, the thought that one should take specific steps to protect oneself against inflation did not cross the minds of many people before the nineteen-seventies. In the years preceding the prolonged inflation that began in 1965, the general public usually anticipated price stability. There was a series of inflationary experiences during the first twenty or so years following World War II, 1946–48, 1950–51, and 1956–58, but none was so prolonged as to replace in the public's mind the existing belief that relative price stability was the norm and inflation was the exception. However, by 1970, there had been five years of uninterrupted inflation, each year being worse than the preceding. By then most of the respondents of consumer surveys, such as the Survey of Consumer Finances conducted by the University of Michigan, expressed an expectation of inflation for the coming year equal to or greater than the existing rate. The widespread expectation of inflation that began about that time has persisted since then, because the actual record of inflation over these years has given the public no basis to expect anything but more of the same. If the sharp slowdown in the rate of inflation that occurred in 1976 continues, this expectation could change in the years ahead, but as of this writing in the spring of 1977 it is safe to say that the public does not believe it has seen the end of this record-long period of inflation.

As measured by the impact of inflation on one's income and wealth, anyone who is able to increase his money income and the market value of his wealth as fast as prices rise is, from this point of view, not hurt by inflation. However, in order to say that such a person does not suffer a loss, his income and wealth increases must result automatically from inflation, e.g., a money wage rate increase that results from a cost of living adjustment, and not for any other reason. If the decrease in a person's real income and real wealth that otherwise would result from inflation is not offset automatically but rather by his working the extra number of hours per week needed to achieve that result, then this person's

[2]Mention should be made of at least one of a number of other developments that may be called inflation but are not covered by our definition. If an almost surely otherwise persistent and appreciable rise in the general price level is prevented during wartime by price controls and rationing, the underlying inflation is described as "suppressed" or "repressed." Although all may agree that there is some amount of suppressed or repressed inflation in such a period, the extent of such inflation is no more measurable than anything else that is not observable. The conventional definition adopted here is limited to observable and so to measurable changes in prices.

real income and real wealth are unchanged. However, he has clearly suffered a loss from inflation in the form of the sacrifice of leisure or in the form of extra real income he would have had from this same sacrifice of leisure under a stable price level. As we look into the effects of inflation on the income and wealth of different groups, the answer to the question of whether a group is worse off or better off as a result of inflation must be viewed in this sense. One obviously cannot conclude that all those families that were better off in absolute terms in 1977 than in 1968 were for this reason not hurt by the decade of inflation. One also cannot deny that some of those families that were worse off in 1977 than in 1968 were, despite this fact, still helped by the same decade of inflation.

We will look at the way that inflation damages the position of some members of society and improves that of others under three headings. First, and most familiar, is the effect of inflation on the distribution of income, second is the effect on the distribution of wealth, and third, and of a different nature, is the effect on the level of income or output produced and the rate of growth of output.[3]

[3]Although we will here look at the effect of inflation on the distribution of income and the distribution of wealth separately, some items in the literature cover both so that we at this point provide a sampling of the literature on the questions of both income and wealth redistribution through inflation: G. L. Bach and A. Ando, "The Redistributional Effects of Inflation," in *Review of Economics and Statistics*, February 1957, pp. 1–13; A. Brimmer, "Inflation and Income Distribution in the United States," in *Review of Economics and Statistics*, February 1971, pp. 37–48; J. Foster," "The Redistributive Effects of Inflation—Questions and Answers," in *Scottish Journal of Political Economy*, February 1976, pp. 73–98; A. M. Maslove and J. L. Rowley, "Inflation and Redistribution," in *Canadian Journal of Economics*, August 1975, pp. 399–409; G. L. Bach and J. B. Stephenson, "Inflation and the Redistribution of Wealth," in *Review of Economics and Statistics*, February 1974, pp. 1–13; E. C. Budd and D. F. Seiders, "The Impact of Inflation on the Distribution of Income and Wealth," in *American Economic Review*, May 1971, pp. 128–38; and S. E. Harris, *The Incidence of Inflation: Or Who Gets Hurt?*, Study Paper No. 7, Study of Employment, Growth, and Price Levels, Joint Economic Committee, U.S. Congress, 1959.

The Effect of Inflation on the Distribution of Income

As everyone knows, those whose money incomes do not rise as fast as the price level suffer a reduction in real income with every rise in the price level. Those fortunate enough to be in the opposite position, of course, enjoy an increase in real income as the price level rises. In trying to answer the question of which groups suffer a loss of income and which enjoy a gain of income, we will first look at a few more or less familiar generalizations bearing on this question. Following this we will look at a bit of empirical evidence.

Inflation and Income Distribution: Who Loses? Who Gains? For most people, wage or salary income makes up practically their total income, so to stay even with inflation requires that their labor earnings per time period for a given amount of work rise at the same rate as the price level. For that portion of the labor market which is nonunionized, there is a tendency for increases in money wage rates to lag behind increases in prices. Because the wage rates in such markets are set by competitive supply and demand conditions, the lag in the upward adjustment of money wage rates will tend to vary inversely with the tightness of the labor market, but typically some lag is to be expected. In the later part of a cyclical upswing as labor markets grow increasingly tight, workers with skills and training may be able to catch up in full. On the other hand, in the case of unionized industries there can be an even longer lag because wage rates are set by contracts, which run for a period of years. Of course, for those contracts that include a full cost of living adjustment (COLA), the lag is relatively short. Beyond this, the wage increase provisions of contracts that do not contain a full COLA are invariably negotiated with allowance for some expected rate of inflation, given the price level experience of the years since 1965. Depending on the terms that are worked out and the actual rate of inflation

that occurs, the allowance for inflation in the compensation of workers under some contracts in some years may turn out to more than cover the increases in the price level.

One sector of the labor force that is generally believed to fall behind is that composed of salaried, white-collar workers. Traditionally, the lag of salary increases behind price increases was especially long for this sector, in particular for state and local government employees including teachers, but this is probably less true today than it once was. Unionization of government employees, including teachers, has grown rapidly, and bargainers for them seek to gain escalator clauses in contracts to protect the membership from inflation as do bargainers for unions representing hourly-rated industrial workers.

Turning to other income shares, a generalization with a history that can be traced back to the nineteenth century is that those persons whose incomes are derived from profits gain from inflation because the prices at which goods sell rise more rapidly than do their costs of production. Under purely competitive conditions, firms are, of course, price-takers and not price-setters. As will be covered in the analysis of the causes of inflation in the following chapter, the occurrence of excess aggregate demand will in the first instance pull up the price level and give rise to an increase in profits. Under purely competitive conditions in the labor markets, the higher price level will lead with some lag to a rise in wage rates as employers, under the inducement of the greater profits, seek to expand output and hire more labor in order to do so. As long as there is such a lag, profits gain at the expense of wages, and those whose incomes are derived from profits gain through inflation. However, the major part of the economy's total output is not sold in purely competitive markets and a large part of the labor force does not have its wage rates determined in anything like purely competitive markets, so we must look also at the area of imperfect competition.

In this area, one deals with firms that have a degree of control over price and accordingly have a pricing policy. If such firms always set prices according to some form of cost-plus pricing, e.g., gross-margin pricing, full-cost pricing, or rate-of-return pricing, one would expect, in general, some lag in the adjustment of prices to costs, except for increases in costs like contractual labor costs which are known in advance. Strict adherence to a cost-plus pricing policy would not improve the relative position of those whose incomes are derived from profits because prices would rise only in line with costs. However, there is also a purely opportunistic element in pricing which appears under certain conditions, one such condition being inflation. Given a surge in aggregate demand which makes possible the sale of capacity output at higher prices than before, some firms may be expected to take advantage of the opportunity by posting increases which are unrelated to costs, competitive conditions, market share, and other considerations that ordinarily enter into the firm's pricing policy. Thus, in a period of inflation, one has the generalization that those whose incomes are derived from profits gain from inflation.

While this is a generalization that has become commonplace over a long span of years, a look at some figures for the last twenty-five years does not lend much support to it for that particular time period. The data in Table 22-2, which we will turn to shortly, suggest that relative to other shares in the national income, profits have fallen behind in years of inflation and surged ahead in years of relative price stability.

Another generalization, one that is hardly open to dispute, is that the recipients of income derived from interest and rents whose amounts are fixed by long-term contracts are hurt by inflation. This generalization, of course, pertains only to income which takes these forms and is not valid for even these forms if the inflation has been correctly anticipated and allowance therefore included in the terms of the

contracts. However, all too often inflation is not anticipated. For example, people who bought long-term Aaa corporate bonds in 1964, a time when a continuation of the price stability that had been maintained since 1958 was expected, received a return of about 4.4 percent or $44 per $1000 on their funds in that year. Ten years later consumer prices were up about 60 percent, but the amount of interest received per $1,000 face value of these bonds was not up one penny. There is no question that the income a person derives from such assets did not keep up with inflation; this income in real terms had decreased by a full two-thirds over the ten year period. Of course, as the lender lost real income, the borrower gained real income, since he paid the interest owed in depreciated dollars. In general, under these conditions there occurs a redistribution of income from creditors to debtors; one group loses income and the other gains income.

If the inflation continues, the redistribution also continues year after year until the contract reaches maturity date.[4] It is, of course, true that once a long-term contract, say a bond, matures, the terms of a new one entered into will reflect the higher interest rates that may then be in effect. They will also reflect the further increases that may result from expected inflation, but nobody can do any more than

[4]The holder of a long-term bond issued by the federal government or a major corporation can, of course, sell it in the market at any time. However, if the market has correctly anticipated the inflation over the remaining life of the bond, the seller will not avoid further loss by selling but will take the balance of his "beating" in a lump sum instead of in small doses over the years remaining to maturity.

TABLE 22-2

Absolute changes in the percentage shares of national income by type of income during years of relatively large increases and years of decline or relatively small increases in the Consumer Price Index, 1949–75

			ABSOLUTE CHANGE IN PERCENTAGE SHARE		
(1)	(2)	(3)	(4)	(5)	(6)
TYPE OF INCOME	PERCENTAGE SHARE 1949	PERCENTAGE SHARE, 1975	TWENTY-SIX YEARS, 1949–75	THIRTEEN YEARS OF RELATIVELY LARGE INCREASES[1]	THIRTEEN YEARS OF DECLINE OR RELATIVELY SMALL INCREASES[2]
Compensation of employees	66.4	76.9	+10.5	+8.1	+2.4
Wages and salaries	63.3	66.8	+ 3.5	+3.7	−0.2
Corporate profits[3]	12.7	7.6	− 5.1	−8.0	+2.9
Proprietors' income[3]	17.0	7.5	− 9.5	−2.3	−7.2
Rental income of persons[4]	2.8	1.9	− 0.9	−1.1	+0.2
Net interest income	1.0	6.2	+ 5.2	+3.4	+1.8

[1]1950–51, 1956–58, and 1965–75
[2]1949–50, 1951–56, and 1958–65
[3]Includes inventory valuation and capital consumption adjustments.
[4]Includes capital consumption adjustment.

SOURCE: U.S. Department of Commerce, *Survey of Current Business*, July 1976 and Part II, January 1976.

guess where interest rates will be five, ten, or twenty years from now, however many pretend they can do more than this. As long as people enter into long-term contracts that do not contain escalator clauses to periodically adjust the interest rate paid to exactly offset declines in the purchasing power of the monetary unit, some will inevitably lose income and some will inevitably gain income in this way as a result of inflation.

Last but not least, the group that traditionally has been identified as the one on which the ravages of inflation fall most heavily is composed of those retired and aged persons whose incomes are almost entirely derived from private pensions and old-age social security benefits. The problem remains, but it is less serious than it was some years ago. Social security beneficiaries have been protected against inflation since 1972 by an escalator formula that automatically increases the size of benefit payments to compensate for increases in consumer prices. Most private pension funds continue to provide fixed dollar amounts, but it is common to base the fixed dollar amount of benefits on the employee's earnings over the last few years of employment. To the extent that an employee's earnings kept up with inflation until the time of retirement, the pension at least allows for whatever inflation had occurred as of the time of retirement. Of course, a few years of rapid inflation following retirement can wipe out a sizable part of such fixed dollar amount pensions, an experience with which people who retired around 1970 are all too familiar.

Inflation and Income Distribution: Some Empirical Evidence Generalizations such as those above for different classes of income recipients are informative, but they say nothing specifically about how much the recipients of one kind of income gain or lose relative to the recipients of other kinds of income as a result of inflation. There is a great deal of data on the changes in the distribution of income that occur over time, but these changes come about for many reasons, inflation being only one. To what degree are the observed changes in the share of national income made up of wages and salaries, corporate profits, self-employment earnings and the like the result of inflation? Although what it reveals is anything but conclusive, there is a simple technique that may be employed here to shed some light on these questions. For example, if one finds that the share of the national income in the form of wages and salaries regularly increases during relatively rapid inflation but decreases or at least does not increase during other periods, this is some evidence that it is inflation that has caused this share of income to have increased as it did. If no such relationship is found, then there is some evidence that inflation does not affect the relative shares of such groups.

Columns 2 and 3 of Table 22-2 show the percentage of the national income accounted for in 1949 and 1975 by the various shares listed in Column 1. Column 4 shows the absolute changes in the percentage shares for the twenty-six year period. As measured by the consumer price index, there was a rise in the price level in each of these years with the single exception of 1954 to 1955, but the rate of price advance showed wide differences from less than 0.5 percent for 1953–54 to 11 percent for 1973–74. The twenty-six years have been divided into the thirteen during which the inflation rate was relatively large, 1950–51, 1956–58, 1965–75, and the thirteen during which the rate was negative or relatively small, 1949–50, 1951–56, and 1958–65.[5] Column 5 shows the cumulative change in the percentage share that occured during the thirteen years of relatively rapid inflation and Column 6 shows the same for the other thirteen years.

As shown by Column 4, over the full period

[5]If the GNP deflator were used instead of the consumer price index, there would be only one change in the classification: 1955–56 would shift into the thirteen years of relatively large increases and 1957–58 would shift into the thirteen years of relatively small increases.

the share of national income in the form of compensation of employees increased by 10.5 percentage points. Compensation of employees is the sum of wages and salaries and supplements like employer contributions for social security and for private pension and welfare funds, which became increasingly important over these years. The share of national income made up of wages and salaries alone increased by 3.5 percentage points over the full period. If one now looks at the figures in Columns 5 and 6 one finds that of the 3.5-percentage point increase in the wages and salaries share, 3.7 percentage points occurred during the years of most rapid inflation and −0.2 points during the other years. The same relationship is found but to a somewhat smaller degree for the compensation of employees share. For the overall period, these results suggest that the redistribution of income effected by relatively large rises in prices has not fallen on the labor share but quite the contrary. A much larger part of the increase in labor's share of the national income pie has occurred during years of relatively rapid rises in the price level than during the other years. In the case of wages and salaries alone, there was actually a minor decrease in the share during the other years.

The figures in the table suggest that over the 1949–75 period the net income of corporations suffered the largest adverse redistributional effect from relatively rapid inflation. For the full period the share of national income composed of corporate profits decreased by 5.1 percentage points. The thirteen years of relatively rapid inflation show a change in this percentage share of −8.0 percentage points; the other thirteen years show a change of +2.9 percentage points. While the share made up of corporate profits thus shrank for the period as a whole, during the years of relatively small price advances the profit share managed to post a gain. From this evidence, there is some indication that over the past quarter century labor has been much better able to protect itself against the ravages of inflation than have the corporations.

While these results suggest that, in terms of its effect on the distribution of income, inflation is good for labor and bad for the owners of corporations, they shed no direct light on the question usually asked in this area: Does the income redistribution brought about by inflation benefit the "rich" at the expense of the "poor"? Although the figures in Table 22-2 give some support to the conclusion that "workers" gain at the expense of "stockholders" during inflation, this is not the same as saying that the poor benefit at the expense of the rich. The labor share and the share of the poor are not the same, but it is still safe to say that the poor and moderate income households gain more from an increase in the wages and salaries share than from an increase in the corporate profits share. From the way that these shares have changed during years of varying inflation rates over the period covered in Table 22-2, one may find some support for the conclusion that the shifts in the distribution of income that occur during years of inflation are more likely to be against the rich and higher income households and in favor of the poor and moderate income households than the other way around.

Additional support for this kind of conclusion is found in the data of Table 22-3. This table covers the same overall time period and is set up in the same way as Table 22-2. However, instead of showing the percentage of total income accounted for by each type of income, it shows the percentage of income received by each fifth of all families. Starting with the fifth of all families with the lowest incomes, one finds a striking result: Of the 0.9-percentage point increase in this quintile's share, almost all of it, 0.8 percentage points, came during the years of relatively rapid inflation. The same kind of result but in reverse is found for the fourth quintile: All of the 0.6-percentage point increase in its share occurred during years in which the price level declined or increased relatively slowly. This suggests that the lowest-income quintile improved its share the most during years of relatively rapid inflation, while the opposite is true for the fourth quintile. The top quintile shows

a 1.7-percentage point decrease in its share, but the decrease is distributed almost evenly between the two subperiods. Perhaps this says that the highest-income families are better able than others to protect themselves against rapid inflation, although they are unable to prevent some decrease in their share over the years.

To the extent that one can sum up for the table as a whole, there is a clear indication that the income shares of the lowest two fifths, especially that of the lowest fifth, fare better during the years of relatively rapid increases in the price level than during the other years. For the middle and fourth fifths, the indication is quite clearly the opposite: Their income shares did much better during the years of price decline or of relatively small increases.

The Effect of Inflation on the Distribution of Wealth

A household whose money income and wealth both rise faster than the price level is a clearcut winner in the inflation race and one whose money income and wealth both rise slower than the price level is an equally clearcut loser. However, one household's money income may go up faster than the price level while the value of its wealth goes up slower or even goes down; another's money income may go up slower than the price level or even decline while the value of its wealth goes up faster than the price level. In these cases, more information is needed to say whether or not the households in question are hurt by the inflation, because for each there is a gain in one direction and a loss in the other direction.

Our concern here is only with the way that wealth positions of persons or households are affected by inflation. As in the preceding section on income, we will first note a few generalizations bearing on the question of whether a household's wealth will expand or shrink as a result of inflation. This will be followed by a brief look at some empirical evidence bearing on the same question.

TABLE 22-3

Absolute changes in the percentage shares of aggregate income received by each fifth of families during years of relatively large increases and during years of decline or relatively small increases in the Consumer Price Index, 1949–75

			ABSOLUTE CHANGE IN PERCENTAGE SHARE		
(1)	(2) PERCENTAGE SHARE, 1949	(3) PERCENTAGE SHARE, 1975	(4) TWENTY-SIX YEARS, 1949–75	(5) THIRTEEN YEARS OF RELATIVELY LARGE INCREASES[1]	(6) THIRTEEN YEARS OF DECLINE OR RELATIVELY SMALL INCREASES[2]
Lowest Fifth	4.5	5.4	+0.9	+0.8	+0.1
Second Fifth	11.9	11.8	−0.1	+0.2	−0.3
Middle Fifth	17.3	17.6	+0.3	−0.5	+0.8
Fourth Fifth	23.5	24.1	+0.6	0.0	+0.6
Highest Fifth	42.8	41.1	−1.7	−0.8	−0.9

[1]1950–51, 1956–58, and 1965–75
[2]1949–50, 1951–56, and 1958–65

SOURCE: U.S. Department of Commerce, Bureau of the Census, *Current Population Reports, Consumer Income, Money Income in 1974 of Families and Persons in the United States,* 1975, Table 22, p. 37, and *Current Population Reports, Consumer Income, Money Income and Poverty Status of Families and Persons in the United States, 1975 and 1974 Revisions,* September 1976, Table 5, p. 17.

Inflation and Wealth Distribution: Who Loses? Who Gains? The way that the wealth positions of households are affected by inflation depends on the way that inflation affects the money value of the particular assets owned by each household and the money value of the debts owed by each household. The debts found on any household's balance sheet are, with negligible exceptions, fixed in dollar terms. The balance it owes on the house mortgage, on the car, on bank loans and the like are all claims against it that do not become larger or smaller in money terms as prices rise or fall. They are fixed claim debts. Among its assets, however, are some that are fixed claims it holds against others and some that are variable-price assets, the latter including all kinds of physical assets like houses, land, automobiles, and other personal property, and financial assets like shares of stock. Among holdings of fixed-claim assets are savings and demand deposits in banks, share accounts in savings and loan associations, and bonds issued by governmental units and private corporations. Although not a claim like the others, currency is also a fixed-claim asset because it is fixed in dollar terms like the other items noted here.

As a household's wealth or net worth is the difference between the value of its assets and debts, its wealth will be adversely or favorably affected by inflation depending primarily on how its assets are divided between variable-price assets and fixed-claim assets and on how large its fixed-claim assets are relative to its debts, which are all fixed claim. Some households will be adversely affected and others favorably affected, or, in other words, inflation will result in a transfer of wealth from some to others. Because the process by which inflation redistributes wealth is probably less familiar and is also less obvious than the process by which it redistributes income, a simple numerical illustration may be helpful. Table 22-4 shows hypothetical balance sheets for a negative, a moderate, and a high net-worth household. The makeup of the three balance sheets shown

TABLE 22-4

Hypothetical balance sheets of negative net-worth, moderate net-worth and high net-worth households before inflation

	Negative Net-Worth Household		
Fixed claim assets	$ 100	Debts	$ 3,500
Variable price assets	2,900	Net worth	−500
Total assets	$ 3,000	Debts plus net worth	$ 3,000
	Moderate-Net-Worth Household		
Fixed claim assets	$ 1,000	Debts	$ 10,000
Variable price assets	14,000	Net worth	+5,000
Total assets	$ 15,000	Debts plus net worth	$ 15,000
	High-Net-Worth Household		
Fixed claim assets	$ 30,000	Debts	$ 15,000
Variable price assets	70,000	Net worth	+85,000
Total assets	$100,000	Debts plus net worth	$100,000

is in rough accordance with that of the typical household at each of these net worth levels.

Starting with these balance sheets, what will be the effect on the relative position of each of these households if there is a doubling of the price level? A major difficulty in arriving at an answer is that different variable-price assets rise at different rates as prices in general rise. For example, over recent years the value of residential property has risen much more rapidly than the general price level, but the value of stock shares has not. However, we must here simplify by assuming that all variable price assets rise at the same rate and that that rate is equal to the rate of increase in the general price level. This means that the *real* value of each household's original holdings of variable-price assets remains unchanged. On the other hand, given a doubling of the price level, the *real* value of each household's fixed-claim assets and debts is cut in half.

Table 22-5 shows the balance sheets of Table 22-4 adjusted for the real changes in the assets and debts of each that occur due to the inflation. The negative net-worth household shows a rise in real net worth from minus $500 to +$1,200. Inflation has wiped out its insolvency! In the same way one finds that the real net worth of the moderate net-worth household has risen from $5,000 to $9,500 and that of the high net-worth household has fallen from $85,000 to $77,500.

The negative net-worth households, which are in general the households that are also poor in terms of income, are among the major beneficiaries of inflation. On the average, they do not show large amounts of debt—they are poor credit risks and find it difficult to borrow—but they also show even less of fixed claim assets. What little assets they have are primarily in their household possessions and an automobile. It is this composition of assets and the low ratio of debt to fixed claim assets that brings them out ahead in a period of inflation. From this it is easy to appreciate why, on the average, they show no concern about inflation so far as its effect on their net worth is concerned.

TABLE 22-5
Hypothetical balance sheets of negative net-worth, moderate net-worth, and high net-worth households after inflation

	Negative-Net-Worth Household		
Fixed claim assets	$ 50	Debts	$ 1,750
Variable price assets	2,900	Net worth	+1,200
Total assets	$2,950	Debts plus net worth	$2,950
	Moderate-Net-Worth Household		
Fixed claim assets	$ 500	Debts	5,000
Variable price assets	14,000	Net worth	+9,500
Total assets	$14,500	Debts plus net worth	$ 14,500
	High-Net-Worth Household		
Fixed claim assets	$15,000	Debts	$ 7,500
Variable price assets	70,000	Net worth	+77,500
Total assets	$85,000	Debts plus net worth	$ 85,000

The net worth of the moderate net-worth household shows a gain of $4,500 in real terms. The major factor here is that households at this net worth level typically show large amounts of debt relative to fixed-claim assets. Many will own their own homes and a major part of their debt will be a mortgage balance owed on it. At the same time that they will have some fixed claim assets in the form of bank accounts and the like, they will hold by far the major part of their assets in the variable-price form. Their single largest asset in most cases will be their home.

For the high net-worth household, Tables 22-4 and 22-5 show a decline of $7,500 in its real net worth. It is a loser. Relative to its net worth, the amount of its debt typically will be small. It ordinarily does not find it necessary to borrow to cover purchases that lower net-worth families can only cover by borrowing. On the asset side, it will tend to hold a sizable part of its total assets in fixed-claim form like bonds. For one thing, if a *very high net-worth* household is also a *very high income household*, there is a great advantage in holding state and local government securities as the interest on these is not subject to federal income tax.

Given the assumptions made, the changes in the real net worth of the three illustrative households will be as shown. However, there is an important qualification to this not covered by the illustration: The indicated gains and losses are based on the assumption that the inflation was unanticipated. If at any time inflation is generally anticipated, households and others will make longer-term loans and buy bonds and other fixed-claim instruments only if they are able to secure an interest rate on these high enough to compensate, at least in part, for the fact that the dollars they will get back when the loans come due and the bonds mature will be dollars of smaller real value. Households who seek to obtain loans and others who seek to issue bonds and other fixed-claim instruments will be willing to pay higher interest rates as they expect to repay the amounts borrowed in dollars of smaller real value. However, as noted earlier, although practically everyone had come to anticipate more inflation by the beginning of the seventies, not everyone anticipated the same rate and what that rate turns out to be for any time period can only be known after the fact. Therefore, unless a lender arranges with a borrower that the principal amount to be repaid at the maturity of the contract shall be adjusted by an agreed-upon price index that returns to the lender the same amount of real purchasing power supplied at the beginning of the contract, there will still be redistribution of wealth between debtors and creditors to the degree that the actual rate of inflation over the period covered by contracts of different maturities differs from the anticipated rate for that maturity or the rate reflected in the terms of the contract. However unlikely this may be, there is even the following possibility under certain conditions: The consensus view of the anticipated rate of inflation turns out to be so much higher than the actual rate of inflation that the redistribution of wealth between creditors and debtors turns out to be larger than it would have been if inflation had not been anticipated.

A final note on our illustration in Tables 22-4 and 22-5. The figures there have been chosen to be consistent with the fact that all households combined occupy a net creditor position in the economy. The sum of fixed-claim assets of the three households in Table 22-4 is $31,100 and the sum of debts is $28,500 for a net-creditor position of $2,600. Table 22-5 shows that the net-creditor position in real terms is cut in half to $1,300 after prices double. While we found a decrease in real net worth for the high-net-worth household and an increase for the other two as a result of the inflation, for all three combined we find a decrease in real net worth, as would be true for all households in the actual economy. In the redistribution of wealth effected by inflation, households as a group at this stage appear to lose to business firms and to government units, which as a group are in a net debtor position.

Inflation and Wealth Distribution: Some Empirical Evidence Our illustration with its limitation to three representative households with different net worths has aimed only at identifying, in general, the process by which the distribution of wealth is changed through inflation. To arrive at specific estimates of the extent of such changes that occur among households is clearly another matter and one that has not received as much study as its importance deserves. Table 22-6 gives the estimates reached in one study.

Households are ranked by size of initial (before inflation) net worth and grouped into "quantile" classes of various sizes. In Column 1, the first class, 1–10, includes the 10 percent of households with the smallest net worth and so forth through the last class, 96–100, which includes the 5 percent of households with the largest net worth. Column 2 shows the mean net worth of the sample of households in each class based on asset holdings and debts as of the end of 1962. At the two ends, the 10 percent of households with the smallest net worth had mean net worth of −$415 (something like the −$500 of the lowest net worth household in our illustration), and the 5 percent of households with the largest net worth had mean net worth of $213,900. Column 3 shows the relative share of the total net worth of all households in each quantile. The bottom 10 percent of all households had a −0.21-percent share and the top 10 percent had a 65.28-percent or almost two-thirds share.

This brings us to the last two columns, which show the estimated changes in the distribution of wealth that result from a simulated 5-percent rate of inflation. Column 4 shows the percent change in shares of the total real net worth of households in each quantile, and column 5 shows the percent change in the mean amount of real net worth for households in each quantile. Column 5 measures the change for each quantile relative to the mean amount for each quantile, whereas Column 4 measures the change for each quantile relative to the mean

TABLE 22-6
Effect of simulated inflation rate of 5 percent on value of and shares in real net worth for quantile groups

(1) SIZE OF QUANTILE	(2) MEAN NET WORTH OF QUANTILE	(3) RELATIVE SHARE	(4) PERCENT CHANGE IN SHARES	(5) PERCENT CHANGE IN REAL VALUE
1–10	$ −415	−0.21	10.55	10.48
11–30	248	0.25	11.05	11.10
31–40	2,000	1.00	5.03	5.08
41–50	4,503	2.24	3.69	3.74
51–60	7,479	3.73	1.52	1.58
61–70	11,328	5.65	0.75	0.81
71–80	16,901	8.43	0.19	0.24
81–90	27,767	13.84	−0.41	−0.35
91–95	47,150	11.75	−0.41	−0.36
96–100	213,900	53.53	−0.36	−0.30
All Units	20,050	100.00		

SOURCE: Derived from Table 3 of E. C. Budd and D. F. Seiders, "The Impact of Inflation on the Distribution of Income and Wealth," in *American Economic Review*, May 1971, p. 134.

for the distribution as a whole. The results in the two columns are seen to be very similar and we may limit our attention to one column. We see, for example, in Column 4 that the share of the lowest 10 percent of households increases by 10.55 percent. Since it was −0.21 percent, it rises to −0.19 percent. At the other end, the share of the top 5 percent which was 53.53 percent is reduced by 0.36 percent or to 53.34 percent of the total. For the distribution as a whole, the results suggest a fairly consistent shift to less inequality, although the groups just below the very top appear to lose a little more than the top group. The degree of the shift in real net worth between groups is seen to be very small, but then a five-percent inflation rate is not, in terms of recent world experience, a very rapid rate of inflation. It is interesting to find that there is, according to this study, a perceptible shift toward less inequality in the distribution of wealth as a result of this moderate rate of inflation. Although the figures in Tables 22-4 and 22-5 are only hypothetical, they give some idea of what a 100-percent rate of inflation can do.

In its focus on the way that inflation causes shifts in the distribution of real net worth among households, Table 22-6 provides some answer to the kind of question people usually ask in connection with the redistribution of wealth. Do those households already with a lot get still more? Do those already with little give up some of that little? Do those already with less than none end up with even less than that? A different kind of question that may be asked in the present connection is whether households taken all together win or lose relative to other sectors of the economy. If households on balance are in a net creditor position, i.e., the sum of their fixed claim assets exceeds the sum of their debts, it would appear that they lose wealth as a result of inflation. The decrease in the real value of their assets is greater than the shrinkage in the real amount they owe. If they are in a net debtor position, the opposite conclusion appears to follow. By finding whether each other sector is in a net creditor or debtor position, one may in the same way identify whether each is a loser or a gainer as a result of inflation. A much finer classification is possible, but we will limit attention to only three sectors: households, governments including federal, state, and local, and businesses including corporate and noncorporate, financial and nonfinancial.

Households as a group are in a net creditor position. As of the end of 1973, the figures for all households were approximately $1,469 billion in fixed-claim assets and $661 billion in debts for a net creditor position of $808 billion.[6] Although the dollar amounts during the nineteen-thirties were small in comparison with these, households have maintained a net creditor position over the years since that time. Government units showed fixed-claim assets of about $201 billion and debts of about $611 billion for a net debtor position of about $410 billion as of the end of 1973. As of the same date, all businesses combined owned about $2,371 billion in fixed-claim assets and had about $2,805 billion in debts for a net debtor position of about $434 billion, similar in size to that of all government units combined.

To see the magnitude of the shifts in wealth that inflation may produce among sectors with net creditor and debtor figures like these, take, for example, the 12-percent rate of inflation from December 1973 to December 1974. Other things being equal, the $1,469 billion of fixed-claim assets held by households at the end of 1973 would have been worth $1,312 billion ($1,469 billion/1.12) at the end of 1974 for a real loss of about $157 billion. The $661 billion of debts would have shrunk to $590 billion ($661 billion/1.12) for a real gain of about $71 billion. The net loss would have been $86 bil-

[6] These and the following figures are derived from the flow of funds accounts, *Federal Reserve Bulletin*, October 1974, Table 5, p. A59–14. Under households' fixed-claim assets, only $219 billion of pension fund reserves of $307 billion were included, the balance being deducted as the amount held in corporate shares and thus in variable-price assets.

lion or say in the range of $80 to $90 billion. If one were to make corresponding approximations over the thirty-five years of almost uninterrupted inflation since World War II, the loss of wealth to households through inflation would probably approach $1 trillion measured in dollars of today's purchasing power. Government and business as net debtor sectors would have gained this tremendous sum as households lost it. Most of it would have been transferred to government for it has only been over the last decade or so that all businesses combined have shown a large net debtor position.

Even in a huge economy like that of the United States, a transfer of wealth on a scale anything like this is still staggering. However, there are several qualifications that make the shift less massive than it at first appears to be. One was noted earlier: To the degree that inflation is anticipated, creditors obtain some protection against loss through higher interest rates. The interest rates lenders were able to obtain in 1974 on new loans would surely not have been at the record levels reached in that year in the absence of the then raging inflation and the widespread expectation that it would not slow to a reasonable rate for some time to come. If the $80–$90 billion range for the wealth transfer from households in 1974 is reasonably correct without this allowance, then the appropriate range with this taken into account is something appreciably less, though how much less is very difficult to estimate even in rough terms.

A second qualification arises from the fact that the transfer of wealth from households to businesses and government units is not a transfer in the full sense of the word. From a superficial view, there is no question but that such a transfer occurs, but one must look below the surface to see what particular groups of households in turn benefit from the wealth transferred from households as a group to businesses and to government units. Whatever the indicated transfer to businesses may be, the inanimate business units themselves do not enjoy the gain; it is primarily the households who own the businesses who are the beneficiaries. Businesses which maintain the same rate of output will find, on the average, that a 10-percent rise in the price index will be accompanied by a 10-percent rise in their receipts and a 10-percent rise in their labor costs, raw materials costs, and the like. However, they will find no corresponding increase in the number of dollars needed to meet the interest cost on their existing debts. Therefore, more dollars remain after meeting all expenses and these extra dollars pass to the owners of the businesses.

What we thus find is first a transfer of wealth from households as a group, net creditors, to businesses as a group, net debtors, but then a transfer from businesses back to households with different households participating in this according to their ownership interest in business. As the owners of businesses are not evenly distributed by wealth or net worth classes but are found mostly in the higher net worth classes, the transfer of wealth that occurs in this complicated way tends to be from lower net worth classes to higher net worth classes. Note that this works in the opposite direction to the shift in wealth shown as a result of the process examined in connection with Table 22-6.

Like businesses, government units are net debtors and inflation initially produces a transfer of wealth from households to government units, but as with businesses the process does not end there. Who benefits from the transfer to government is not inanimate government but people or households, and the primary beneficiaries may be taxpaying households. Suppose as we did above that a 10-percent rise in the price index occurs. With no other changes, government tax receipts will rise by roughly 10 percent. We may assume that the inflation is accompanied by an approximately equal rate of increase in the salaries of government employees and in the cost of supplies, equipment, and everything government units buy. However, government units will not face an equal increase in their outlays for interest on exist-

ing debts. Although government receipts and spending have both risen in money terms, in real terms spending will have decreased and the real burden on taxpayers will also have decreased. One finds in this way a transfer back to households. This will be distributed among households by net worth class as taxes are distributed by net worth class.[7] As the top fifth of households so classified pay much more per household in taxes then does the bottom fifth, it appears that the overall process works toward a transfer of wealth from lower net worth classes to higher net worth classes. This is the same result indicated above for the transfer to and back from businesses. And again it appears to work in the opposite direction to the shift shown as a result of the process examined in connection with Table 22-6.

We have now reviewed in connection with Table 22-6 the way that inflation may work directly to bring about shifts in wealth among households classified by net worth and the way that it may do this indirectly by shifts of wealth from households to business and government. Whether the end product is greater or less inequality in the distribution of wealth among households is not a conclusion that can be firmly drawn. Furthermore, to the extent that any conclusion can be reached, one would not expect that conclusion to be the same in all inflations. However, if one insists on a fairly firm conclusion applicable in all cases, perhaps one must settle for the statement found in every introductory economics textbook's list of the consequences or effects of inflation: Debtors gain from inflation and creditors suffer from inflation. But while this is surely correct as a general proposition, we have here gone far enough to see some of the complications that lie behind it and why one cannot unqualifiedly conclude from it that every inflation redistributes wealth from creditors to debtors.

[7]This is not the only possible outcome. Government may increase its spending in money terms by whatever amount is needed to maintain real spending unchanged. In this event, there will be no reduction in the real burden on taxpayers, and the transfer back to households will be distributed in a different way.

The Effect of Inflation on Output Employment and the Growth Rate

In an economy operating below full employment, will a spurt of inflation get some of the unemployed back to work? Will it mean that next year's output will be closer to the economy's potential than was this year's? For some people, the question of whether they personally gain or lose from inflation is answered right here. If they would not have a job but for inflation, they clearly gain in this way from inflation, regardless of whether they gain or lose in terms of the inflation-caused redistribution of income and wealth. Another question is how inflation affects the economy's real wealth. Does it speed or slow the rate at which the economy builds up its stock of real assets? If it speeds it, wealth grows faster than it otherwise would and the potential rate of growth of output becomes greater than it otherwise would be. The first part of this section looks at the question of inflation's impact on output and employment in the short run and the second part looks at its impact on the accumulation of capital and the rate of long-run growth.

Inflation's Impact on Output and Employment in the Short Run For an economy producing below potential, many economists have long maintained that inflation of the creeping or crawling variety will have a tonic effect on output and employment. Here again it is essential to draw the distinction between anticipated and unanticipated inflation, because the alleged beneficial effects in large part disappear if inflation is anticipated. In the event of unanticipated inflation, prices rise faster than money wage rates, and the resulting reduction in the real wage rate gives business the profit incentive to hire more workers and expand output. The result is that a rise in the inflation rate, if unanticipated, may lead to a reduction in the unemployment rate. This line of argument will be developed in detail in the final section of the following chapter. It is sufficient to note here that, in general, an expansionary effect on em-

ployment and therefore on output, at least of a temporary nature, is to be expected from unanticipated inflation.

An increase in output results not only from getting the unemployed back to work but also from reallocating labor and other resources to get the maximum output these resources are capable of producing. While inflation may lead to an increase in output by reducing unemployment, there is an argument that inflation also tends to prevent the optimum allocation of resources and in this way tends to hold down output. The industries that have the strongest incentive to expand during a period of inflation will be those whose prices rise most rapidly relative to costs, but there will be some whose prices are restrained from rising, despite the inflation, and in which little expansion or possibly even contraction may result. A couple of examples will make the point. Although electric utilities in some states now get automatic rate adjustments to cover changes in their fuel costs, they can only get rate increases to cover inflation of other costs after a delay whose length varies with the time taken by state regulatory commissions to act. Prices charged by some industries, e.g., coal company sales to large industrial buyers, are on a contract basis and the inability of firms in such industries to adjust prices may mean losses instead of larger profits and contraction rather than expansion of output. Distortions like these in the price system result in resource misallocation. However, the resource allocation effect on output is not necessarily negative; there is a counterargument that inflation works in another way to bring about a better allocation of resources than would occur with stable prices. Because the price system as it exists is one in which most prices are flexible upward but inflexible downward, the only way in practice to get the relative price changes needed to achieve a better allocation of resources is by having some prices rise. This amounts to saying that the only way that the price system can achieve a better allocation of resources is through inflation.

If the net effect of these forces on resource allocation is a better allocation, the resources employed will turn out a greater total output than otherwise. Therefore, whatever increase in output results from the fuller utilization of resources, there will be some additional amount that results from the better allocation of the resources used. In the opposite case in which the forces, on balance, result in a worsening of the allocation of resources, there will be some decrease in output that must be subtracted from the increased output gained through a fuller utilization of resources.

What then is the answer to the question? If the economy is suffering from unemployment, will a moderate inflation cut unemployment below and raise output above what they otherwise would be? All things considered, most economists would probably answer yes, but many would quickly add important qualifications. Perhaps the most important is that, according to some, the reduction in unemployment achieved through the rise in the rate of inflation will only be temporary. The lag of the money wage rate behind prices, which brings about the reduction in unemployment, will be short-lived. Labor comes to anticipate inflation and acts to prevent the reduction of the real wage rate that can result from unanticipated inflation. Again, this is a question that we will deal with more thoroughly in the final section of the following chapter.

Another important qualification is that at the same time that inflation may reduce unemployment and raise output, it creates distortions of various kinds that will culminate in a downturn. A key role in the process is played by changes in inventories. If the consensus opinion is that there will be a continuation or even acceleration of an already high rate of inflation, business persons may seek to build up inventories far in excess of what they need to meet any expected near-term increase in sales. The purpose, naturally, is to secure that large additional profit some time later by selling the goods at the much higher prices expected then. Beyond building inventories, some businesses may also be encouraged by such expectations to expand ex-

penditures for new plant and equipment. The likelihood is that the cost of borrowing and the purchase price of capital goods will be higher next year and the year after. Consumers also participate in the surge of overbuying. They don't maintain stocks of goods for resale as do businesses, but expectations of continuing inflation will lead some to accumulate durables and nonperishable goods.

Depending on the extent of the excesses, the eventual adjustment will produce anything from a minor slowdown to a serious recession. Some adjustment is inevitable in a process whose foundation is predominantly speculation. If the major reason that buyers as a group expand their buying at today's prices is that they expect to be able to sell tomorrow at substantially higher prices, an unsustainable process has been set into motion. Its end may come because of the inability of business persons to finance the amount of inventories they seek to accumulate. The monetary authorities may in due time refuse to permit credit expansion to continue at the rate needed to accomodate business demands. Bank loans become increasingly hard to get and the interest rates that must be paid soar. The profit possibilities in building inventories still higher begin to look less promising and firms under pressure begin to move in the direction of reducing instead of building inventories. If this occurs on any scale, prices may not only stop rising, contrary to what was generally expected, but there may even be some scattered decreases in prices. As the signs of a reverse like this become clear, many businesses become as anxious to reduce inventories as they were a short time earlier to enlarge them. The inventory adjustment that follows may be orderly and cause no more than a slowdown in the economy or it may possibly turn into something more serious.

We noted above that most economists would probably answer yes to the question of whether inflation will raise employment and output and then inserted some qualifications they would add to this answer. With the qualifications taken into account, what we are left with is anything but a conclusive "yes" or "no" answer to the initial question. Only in the case of an extremely rapid inflation can the answer be fairly conclusive. Although such an inflation may at first stimulate production and reduce unemployment, it will likely lead eventually to a more or less violent adjustment with production lower and unemployment higher than before. On the other hand, a short spurt of moderate inflation will stimulate production and raise employment without generating massive overbuying by consumers, excessive inventory accumulation, and the wild speculative activity that come with a prolonged, sharp rise in prices. The economy may adjust to such a moderate inflation with only a ripple instead of a wave; the gain in production and employment, even though only temporary, need not be offset by a later loss. Believing this to be the case, some economists maintain that intermittent spurts of moderate inflation should be deliberately engineered to meet the problem of an economy that has continuing difficulty in maintaining full employment, but to others this sounds like playing with fire. A little inflation at times has a way of growing into a lot.

As is all too apparent, there is no simple "yes" or "no" answer to the question of whether inflation has a beneficial or detrimental effect on employment and output in the short run here under consideration. The answer varies with the degree of inflation, whether it is anticipated or not, and the particular way that the economy happens to respond to it. Generalizations are, to say the least, hard to come by.

Inflation's Impact on the Rate of Long-Run Economic Growth When one observes the rate at which an economy's output grows over the long run, does one find that the rate of growth is higher during those extended periods that are marked by moderate inflation or during those marked by price stability? Looking back to the record of the eighteenth and nineteenth centuries, some economists find a positive relationship between inflation and economic growth in various countries. The driving force in rapid

progress is identified as large and long-lasting profit margins. As we have seen, such margins can be maintained only as long as wages lag substantially behind prices. In today's world it will not take years for labor to remove such a lag, but some argue that in the past such lags at times prevailed over stretches of many years. To the extent that such lags did occur, they made possible massive shifts of resources away from the production of consumer goods for wage earners to the production of that much more of capital goods. This rapid accumulation in turn provided the increase in productive capacity and labor productivity that made possible extended periods of rapid economic growth.[8]

According to this argument, this shift in resources comes about through the effect of inflation on saving. Given an inflation in which wages lag behind prices, the wage share of total income shrinks and the fraction of total income saved expands. This follows from the fact that a larger portion of the total income goes to profits and other non-wage income, the recipients of which are typically upper income groups with a relatively high propensity to save. That a greater saving will occur under the conditions here specified is well established. On the classical assumption that investment expands to absorb whatever volume of saving is forthcoming, the portion of output that takes the form of capital goods then grows in line with the portion of income that goes into saving.

However, as is the case with so many arguments, there is a counter-argument that maintains that inflation may discourage saving and thus slow the accumulation of capital. Although some people see their incomes rise rapidly because they share in the rapid growth of the profits share, they will also see their real net worth shrink to the extent that they are net creditors. In the face of this, there is some question whether inflation is likely to increase the fraction of their income devoted to saving. The more rapid the inflation, the less likely it is to increase the saving fraction. There is no doubt but that inflation at the breakneck speed called hyperinflation discourages saving. However, as long as the lag of wages behind prices continues, it is likely that moderate or even somewhat more than moderate inflations, which are the typical kinds, will still raise saving relative to consumption spending. For one thing, an inflation premium tends to be built into interest rates to give new purchasers of fixed-claim assets some protection against loss. Apart from this, protection is provided by avoiding ordinary fixed-claim assets. Savers may go into an assortment of variable-price assets that provide varying degrees of protection against loss due to inflation.

Barring the special case of hyperinflation, on balance, the crucial factor in whether or not saving is encouraged by inflation is the existence or nonexistence of a wage lag. While there is considerable evidence of such a lag for the earlier part of this century and for centuries preceding, there is little evidence of this in the United States for the 1949–75 period covered by Table 22-2. The corporate profits and noncorporate business shares of national income declined substantially while the employees' compensation share increased substantially over these years of almost uninterrupted inflation. To the extent that the rate of long-run economic growth depends on the rate of capital accumulation, a major basis for the conclusion that inflation promotes rapid economic growth is undermined, given that wages no longer lag during inflation as they apparently did in times past.

There are, of course, many factors other than the rate of capital accumulation that influence an economy's long-run growth rate. As we saw in Chapter 20 on the relative importance as-

[8] On the question of the wage lag, see A. A. Alchian and R. A. Kessel, "The Meaning and Validity of the Inflation-Induced Lag of Wages behind Prices," in *American Economic Review*, March 1960, pp. 43–66, and T. F. Cargill, "An Empirical Investigation of the Wage-Lag Hypothesis," in *American Economic Review*, December 1969, pp. 806–16; and on the question of the wage lag as a factor in the rate of economic growth, see D. Felix, "Profit Inflation and Industrial Growth," *Quarterly Journal of Economics*, August 1956, pp. 441–63, and E. J. Hamilton, "Prices as a Factor in Business Growth," *Journal of Economic History*, December 1952, pp. 325–49.

signed by Denison to different sources of economic growth, advances in knowledge, for example, are judged to have been more important than the growth in the stock of capital over the 1929–69 period. However, whatever the importance of capital accumulation may be, with substantial wage lags a thing of the past in industrialized countries, inflation no longer appears to contribute to growth in the way that it apparently once did. When one compares the rate of inflation and the rate of growth of real gross national product in major industrialized countries over the period since World War II, one finds no clear pattern. For example, West Germany with the lowest rate of inflation has shown one of the highest rates of growth and Japan with the highest rate of inflation has shown by far the highest rate of growth. The United Kingdom with the lowest rate of growth is among the highest in terms of the rate of inflation. There are problems of interpretation here due to the impact of World War II on different countries, but in general the evidence for the period hardly enables one to say that inflation has either speeded or slowed the rate of economic growth in the major industrialized countries during the long period of inflation since World War II.[9]

THE EFFECTS OF INFLATION: A CONCLUDING NOTE

In looking at the broad question of the effects of inflation, we have gone far enough to see that inflation is a pervasive economic process whose effects are felt to some degree by every citizen and in every corner of the economy. Some of the effects are quite certain, at least when stated as broad generalizations, and others are just as much uncertain. We know that inflation brings about a redistribution of income and wealth, but we have also seen that beyond the generalizations are uncertainties as to the dimension and, in the case of wealth, even the direction of the redistribution that occurs. It would perhaps be incorrect to say that we know that inflation causes short-run changes in the economy's output and employment levels and long-run changes in its growth rate, but it would be hard to believe that inflation leaves these unaffected. But whether inflation raises output and employment in the short run without being followed by declines equal to or greater than the increases is a question on which economists disagree. There is probably no general answer; the answer may well vary from one inflation to another. Similarly, in connection with the last of the effects discussed—inflation's impact on the long-run growth rate—it is not clear for today's advanced economies whether inflation makes for a faster or a slower growth rate than would occur with stable prices.

Taking into account as best one can all the effects considered, the question of whether inflation is beneficial or harmful has a ready answer only when it proceeds at a rapid rate. A rapid rate that is unanticipated and to which there is thus no chance to adjust is undoubtedly damaging. It is when inflation proceeds at a mild rate of not more than a few percentage points per year that there is no ready answer to the question posed. Such an inflation may do more good than harm or more harm than good or perhaps in one case the one and in another case the other. In view of this uncertainty, the best policy to many appears to be one aimed at achieving and maintaining a stable price level or a zero rate of inflation. Considering the record since the mid-sixties, such an achievement looks like a remote possibility, but it does not mean that it is not the best ultimate goal toward which policy should aim.

[9]For one study of this question, see M. Paldam, "An Empirical Analysis of the Relationship between Inflation and Economic Growth in 12 Countries, 1950 to 1969," in *Swedish Journal of Economics* (retitled *Scandinavian Journal of Economics* in 1976), December 1973.

chapter 23

Inflation: The Process and Its Causes

Although it is not the only approach to an analysis of the causation of inflation, the one almost invariably followed is that based on the distinction between supply and demand theories of inflation. Or, in the more familiar terminology, the distinction is between demand-pull theory—the price level is pulled up by excess demand—and cost-push theory—the price level is pushed up by the exercise of market power by labor unions and business firms that hold such power.[1] The first part of this chapter examines the process of demand-pull inflation and the second part does the same for cost-push. The final part looks at the relationship between inflation and unemployment, the famous Phillips curve. Until the late sixties, the evidence suggested that there was a stable relationship between the unemployment rate and the inflation rate, and this relationship in the form of the Phillips curve was at the heart of the discussion of inflation during the sixties. Because evidence over more recent years suggests anything but a stable relationship, the Phillips curve is not the center of attraction it once was. However, it is still very much on stage and its present role will be reviewed.

DEMAND-PULL INFLATION

One main branch of inflation theory runs in terms of generalized excess demand, sometimes loosely and not very accurately described as "too much money chasing too few goods."

[1]In part because the distinction between demand-pull and cost-push is imprecise and because the interdependence between aggregate supply and aggregate demand argues against treating them separately, a recent survey of inflation takes a quite different approach. Its approach is based on the distinction between equilibrium, or perfectly anticipated, inflation and disequilibrium, or imperfectly anticipated, inflation. The disequilibrium side of the approach recognizes the important role played by differences between the expectations and realizations of different groups of workers and businesses in the interaction of prices and wages with excess demand, unemployment, and the like. See D. E. W. Laidler and J. M. Parkin, "Inflation—A Survey," *The Economic Journal*, December 1975, pp. 741–809. The last twelve of these pages provide an extensive bibliography of the recent literature on inflation.

According to this theory, the general price level rises because the demand for goods and services exceeds the supply available at existing prices. In terms of the *IS-LM* framework, the excess demand that pulls up the price level may originate with any one or a combination of a number of changes that can shift the *IS* curve or the *LM* curve to the right. Factors that shift the *IS* curve are real factors and those that shift the *LM* curve are monetary factors. Among the real factors are fiscal actions like changes in government spending and taxes and among the monetary factors are changes in the money supply. In the last section of Chapter 17 we saw how such governmental actions of an expansionary kind could be used to shift the *IS* and *LM* curves to the right, and thereby shift the aggregate demand curve to the right, and, in an economy operating below full employment, raise the level of output and employment.

In Figure 17-8, p. 328, through which that analysis was presented, we worked with an aggregate supply curve that slopes upward to the right before becoming perfectly inelastic at the full-employment output level. Therefore, starting below full employment, rightward shifts in the aggregate demand curve not only raised the output level but also the price level, but the question on which we focused was how full employment might be attained. The rise in the price level that occurred in the process was mentioned only in passing. In here focusing specifically on inflation, we will employ the same graphic apparatus presented in Figure 17-8. Unlike in Figure 17-8, we will here assume that the economy is already operating at full employment, so rightward shifts in the aggregate demand curve affect only the price level or are exclusively inflationary in their effect. Rises in the price level that occur under this condition are what Keynes described as pure inflation.

It was noted in the preceding paragraph that the factors that shift the *IS* and *LM* curves are real and monetary factors, respectively. Demand-pull inflation may originate with either, and it is convenient to divide the discussion of demand-pull inflation into two parts according to whether the inflation originates with real or monetary factors.

Demand-Pull Inflation Originating with Real Factors

Among the various factors that may produce a rightward shift in the *IS* curve are an increase in government spending with no change in tax receipts, a decrease in tax receipts with no change in government spending, a downward shift in the saving function, an upward shift in the export function, a downward shift in the import function, and an upward or rightward shift in the MEI or investment function. The essence of the process by which demand-pull inflation may result from any of these factors is brought out by Figure 23-1.

Assuming that the other determinants of the positions of the *IS* and *LM* curves in Part A of Figure 23-1 are given, there will be a set of *IS* and *LM* curves with each curve in the set corresponding to a different price level. As covered in detail in Chapter 17, a decrease in the price level, other things being equal, will shift the *LM* curve to the right as the decrease in the price level is an increase in the real money supply. An increase in the price level will by reducing the money supply have the opposite effect of shifting the *LM* curve to the left. Thus, Figure 23-1 shows three positions for the *LM* curve corresponding to a given nominal money supply but for three different price levels. As the price level rises from P_1 to P_2 to P_3, the corresponding position of the *LM* curve becomes LM_1, LM_2, and LM_3. In the same way, we saw in Chapter 17 that a decrease in the price level will shift the *IS* curve to the right via the Pigou effect, tax and transfer payments effects, and foreign trade effects. There may also be an income redistribution effect working in this direction but that is uncertain; it may work in either direction. An increase in the price level will have the opposite effect of shifting the *IS* curve to the left. Thus, IS_1, IS_2, and IS_3 in Figure 23-1 make

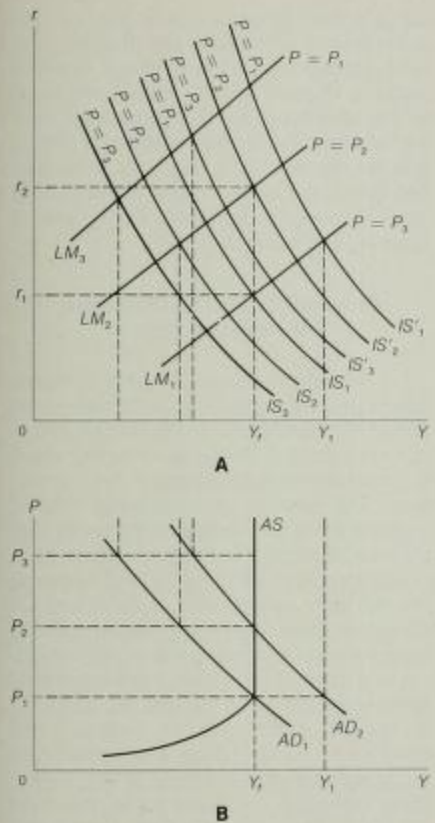


FIGURE 23-1
Demand-pull inflation originating with real factors

up an initial set of *IS* curves corresponding to three different price levels, P_1, P_2, and P_3, the position of the set reflecting the combined influence of the various effects listed.

The intersections of IS_1 and LM_1, IS_2 and LM_2, and IS_3 and LM_3, indicate the amount of goods demanded at price levels of P_1, P_2, and P_3, respectively, and thus provide the information needed to draw the aggregate demand curve labeled AD_1 in Part B of the figure. This *AD* curve intersects the *AS* curve at P_1 and Y_f. Initially there is a full-employment equilibrium at the price level P_1.

One of the real factors that can bring about a demand-pull rise in the price level is an increase in investment demand, i.e., an upward or rightward shift in the MEI or investment function. Suppose there is such a shift in the investment function, which in turn shifts the IS_1 curve to IS_1' and the IS_2 and IS_3 curves to IS_2' and IS_3', respectively. The set of *LM* curves remains as given for there has been no change in the determinants of this set. We now find that *at each price level* saving plus taxes plus imports will equal investment plus government purchases plus exports at a higher *Y* for any *r* or at a higher *r* for any *Y* than before the shift in the investment function.

Following the shift in the set of *IS* curves, the *AD* curve derived from the intersections of the IS_1' and LM_1, IS_2' and LM_2, and IS_3' and LM_3 is AD_2. The new *AD* curve intersects the *AS* curve at P_2, which is the new equilibrium price level. The rise in *P* from P_1 to P_2 is that required to eliminate the excess demand that exists at P_1 after the *IS* curves shift in response to the greater investment spending. At P_1 there is excess demand of $Y_1 - Y_f$. As this excess demand pulls up *P*, there is a decrease in the quantity of goods demanded. With *P* at P_2, the quantity of goods demanded is the amount identified by the intersection of the IS_2' and LM_2 curves in Part A and this quantity is the full-employment quantity. Thus, excess demand is eliminated and there is full-employment equilibrium at P_2. Note also that the process of restoring equilibrium not only requires a rise in the price level but also in the interest rate. As shown in Part A, before the shift in the *IS* curves, equilibrium was found at r_1; after the shift, equilibrium is found at r_2.

The same sequence would follow from any other change that produces a one-time right-

ward shift in the set of *IS* curves. For any such one-time shift, there will be a one-time rightward shift in the aggregate demand curve. Given an original position at the full-employment level of output, that increase in aggregate demand must create an excess demand at the existing price level, which in turn must raise the price level. However, each rise in the price level identifies an *IS* and *LM* curve to the left of those with a lower price level and thus a reduction in the quantity of goods demanded. Equilibrium will be automatically restored by that rise in the price level which reduces the quantity of goods demanded to the amount supplied at full employment.

While this analysis does show how the price level will be pulled up from one particular equilibrium level to another by a specific increase in aggregate demand, it does not deal with the disequilibrium process of inflation, which is a persistent rather than a one time rise in the price level. However, if we return to Figure 23-1, it would appear that such a persistent rise in the price level comes out of this model by merely extending the argument already presented. If there is another and then another rightward shift in the set of *IS* curves, there will be another and then another increase in aggregate demand and in the price level.[2] Such a process may indeed continue up to a point, but the successive increases in the price level must eventually come to an end if the nominal money supply is fixed, an assumption that has been made in connection with Figure 23-1. With Y at Y_f, each rise in P raises the transactions demand for money proportionally with the rise in P. Additional transactions balances can be secured only by drawing them from speculative balances, given that the total nominal money supply is fixed. Once the interest rate has risen to the level at which idle cash balances have been reduced to zero, no additional money is available to generate still higher price levels for the full-employment level of output. Unless the velocity of money were to increase without limit, the rise in the price level must come to an end due to the sheer inability of the existing stock of money to support a higher level of spending and money income.

Demand-Pull Inflation Originating with Monetary Factors

On the monetary side, demand-pull inflation may originate either through a decrease in the demand for money or an increase in the supply of money, but it is the supply of money which is of overwhelming importance in this connection.[3] The essence of demand-pull inflation originating through increases in the money supply may be brought out through Figure 23-2.

This figure is similar to Figure 23-1. Assuming that the other determinants of the *IS* and *LM* curves are given, there is an initial set of *IS* and *LM* curves with each curve in a set corresponding to a different price level. The intersections between the curves in the sets, IS_1 and LM_1, IS_2 and LM_2, and IS_3 and LM_3, yield AD_1, and given the *AS* curve, indicate a full-employment equilibrium with the price level at P_1. There is some increase in the money supply that shifts the LM_1 curve rightward to LM'_1 and the LM_2 and LM_3 curves to LM'_2 and LM'_3, respectively. The set of *IS* curves in this case will remain as given for there has been no change in the determinants of this set. With the shift in the set of *LM* curves, we find that *at each price level* the

[2] Starting from the new equilibrium established at P_2, we would want to identify the positions of a new set of initial curves, IS_2, IS_3, and IS_4, and then, following a change in some real factor, the shift of this set of curves to IS'_2, IS'_3, and IS'_4, a process that parallels that shown in Figure 23-1 for the shift from IS_1, IS_2, and IS_3 to IS'_1, IS'_2, and IS'_3. If the changes in real factors persist, the next set of curves would become IS_3, IS_4, and IS_5, and so forth.

[3] A decrease in the demand for money is not likely to originate an inflation, but it is almost certain to intensify an ongoing inflation that has reached a rapid rate. The greater the rate of inflation, the more costly it becomes to hold money, and the smaller the amount of real balances the public will want to hold at any level of real income and any interest rate. The velocity of money increases, which acts on the price level like an increase in the supply of money.

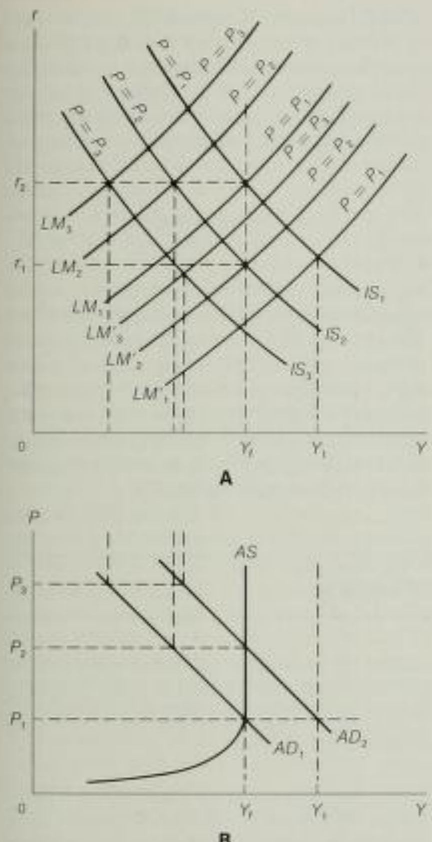


FIGURE 23-2
Demand-pull inflation originating with monetary factors

supply of money will be equal to the demand for money at a higher Y for any r and at a lower r for any Y than before the increase in the supply of money.

With the new set of LM curves, there is a shift in the AD curve from AD_1 to AD_2. At the existing price level of P_1, excess demand equal to $Y_1 - Y_f$ appears. The changes that follow from excess demand are the same for the present case in which the excess demand originates on the monetary side as they were for the case above in which the excess demand originated on the real side. The excess demand will force a rise in the price level of a magnitude needed to eliminate the excess demand. The AD_2 curve intersects AS at P_2 and a rise from P_1 to P_2 is accordingly the rise needed to restore the system to equilibrium. In this case, because the excess demand originates with an increase in the money supply, the interest rate of r_1 at the new equilibrium is lower than the original interest rate of r_2.[4]

Following the analysis above of a one-time shift in the set of IS curves due to a change in some real factor, we asked whether another and yet another such shift in the IS curves could lead to another and yet another rise in the price level. The answer was that rises in the price level would at some point be limited by the fixity of the money supply, although it might be possible to experience a sizable rise in the price level before this limiting point was reached. In now asking the question whether another and yet another rightward shift in the LM curves brought about by another and yet another increase in the money supply can lead to another and yet another rise in the price level, the answer is yes. Unlimited expansion of the money supply will mean an unlimited rise in the price level.

There is no disagreement among economists regarding this last statement: Persistent expansion of the money supply is a sufficient condition for a persistent rise in the price level. Disagreement arises over the importance of real factors in originating and supporting a persistent rise in the price level. Some will agree with Professor Friedman's position that "Inflation is always

[4] This is the way the interest rate is expected to respond in the short run to an increase in the supply of money. However, we will see in a later section that the kind of monetary expansion that leads to a persistent rather than a one-time rise in the price level will be followed by a rise rather than a decline in the interest rate.

and everywhere a monetary phenomenon . . . and can be produced only by a more rapid increase in the quantity of money than in output."[5] This seems to leave little room for inflation to originate with or to be driven by real factors. Others will agree with Professor Hicks who assigns to money at best a supportive role. After noting that "it was true in the old days that inflation was a monetary matter; prices rose because the supply of money was greater than the demand for it," he then adds that "money is now a mere counter, which is supplied by the banking system (or by the government through the banking system) just as it is required."[6] Here the view is that in recent experience it has been real factors that have played the crucial role and that government has been merely permissive or accomodative in not restraining the growth of the money supply as required to prevent the inflation that originates with or is caused by changes in real factors.

The relative role of real and monetary factors in modern demand-pull inflation is thus a matter on which some of the world's leading economists take opposite positions. However, whatever the origin of excess demand, the underlying mechanisms by which that excess demand pulls up the price level is the one illustrated by Figures 23-1 and 23-2, and our purpose here has been the limited one of outlining that mechanism. In these two figures, we assumed that the aggregate supply function was given and thus ruled out any changes on the supply side of the economy. Inflation resulted from excess demand as the aggregate demand curve shifted upward along the fixed aggregate supply curve. As a next step, we turn to the inflation that may result from upward shifts in the aggregate supply function. This branch of inflation theory is called supply, or cost-push, inflation.

COST-PUSH INFLATION

For many years before the fifties, inflation theory ran predominantly in terms of generalized excess demand, the origin of the excess demand being explained either in terms of changes in real factors or monetary factors. During the fifties there occurred a revival of supply or cost theories of inflation. Although then commonly referred to as the "New Inflation," the general notion that price inflation can arise on the supply or cost side is far from new. In a survey of inflationary theory, M. Bronfenbrenner and F. D. Holzman note that

> Cost inflation has been the layman's instinctive explanation of general price increases since the dawn of the monetary system. We know of no inflationary movement that has not been blamed by some people on "profiteers," "speculators," "hoarders," or workers and peasants "living beyond their station."[7]

Of course, we now know much that is "new" about the nature of cost-push inflation as a result of the research of the last several decades that recognizes the structural and institutional changes during this period, but the general notion that inflation can be caused by supply or cost is an old one.

There are two principal causes of supply-side inflation, both of which represent the exercise of market power by specific groups in the econ-

[5]M. Friedman, *The Counter-Revolution in Monetary Theory*, Occasional Paper No. 33, Institute of Economic Affairs, London, 1970, p. 24.

[6]J. R. Hicks, "The Permissive Economy," in *Crisis '75 . . . ?*, Occasional Paper Special, No. 43, Institute of Economic Affairs, London, 1975, p. 17.

[7]M. Bronfenbrenner and F. D. Holzman, "Survey of Inflation Theory," in *American Economic Review*, September 1963, p. 613. An extensive bibliography of the pre-1963 literature on inflation will be found at the end of this article.

omy. One is higher money wages secured by labor unions and the other is higher prices secured by business firms in monopolistic or oligopolistic industries. For purposes of classification, we may call these two principal causes of inflation on the supply side wage-push and profit-push.

Wage-Push Inflation

The belief that rising prices are the result of labor unions' extracting from employers money-wage rate increases greater than the increase in the productivity of labor is one held by a sizable segment of the general public, excluding, of course, the segment made up of labor union members. The idea has a simple plausibility to it: If firms find the labor cost per unit of output rising, they in turn raise their prices to cover the higher cost. A series of increases in wage rates thus leads to a series of increases in prices which is inflation. The wage-push explanation of inflation also has appeal to many because it appears to provide an easy solution to a riddle of recent years. Our last several recessions have exhibited the perplexing phenomenon of pronounced rises in the price level in the face of a rising rate of unemployment, a combination of changes variously described as stagflation and recessionary inflation. While there is surely more to this than can be explained by wage-push, that theory of inf [illegible]
least consistent with [illegible]

that indicated by W_1 in Part C. On the assumption that the money-wage rate does not fall below W_1, the AS_1 curve is derived from this money-wage rate in Part C and from the other parts of the diagram in the way explained in Chapter 13 (see pp. 219–23). The equilibrium price level and output levels are determined in Part D by the intersection of the AD curve and the AS_1 curve. Suppose the IS and LM curves are positioned so as to produce the AD_1 curve in Part D. The result is $P = P_1$, $Y = Y_f$, and $N = N_f$ or the equilibrium level of output is one of full employment of the labor force. Observe that an AD curve to the left of AD_1 would produce an intersection with AS_1 below Y_f and would result in an equilibrium with less than full employment.

With the initial AS_1 and AD_1 curves in place, we can turn to the process by which increases in the money-wage rate push up the price level. Assume that there is an increase in the money-wage rate that results entirely from the exploitation of the market strength of labor unions and in no part from increased productivity of labor or increased demand for labor. This change appears in part C as a shift in the W curve from, say, W_1 to W_2. With the curves in Parts A and B as given, the AS curve derived in Part D is now AS_2. The increase in W from W_1 to W_2 has pushed AS upward from AS_1 to AS_2. The price level at which each possible level of output will be supplied increases proportionally with the

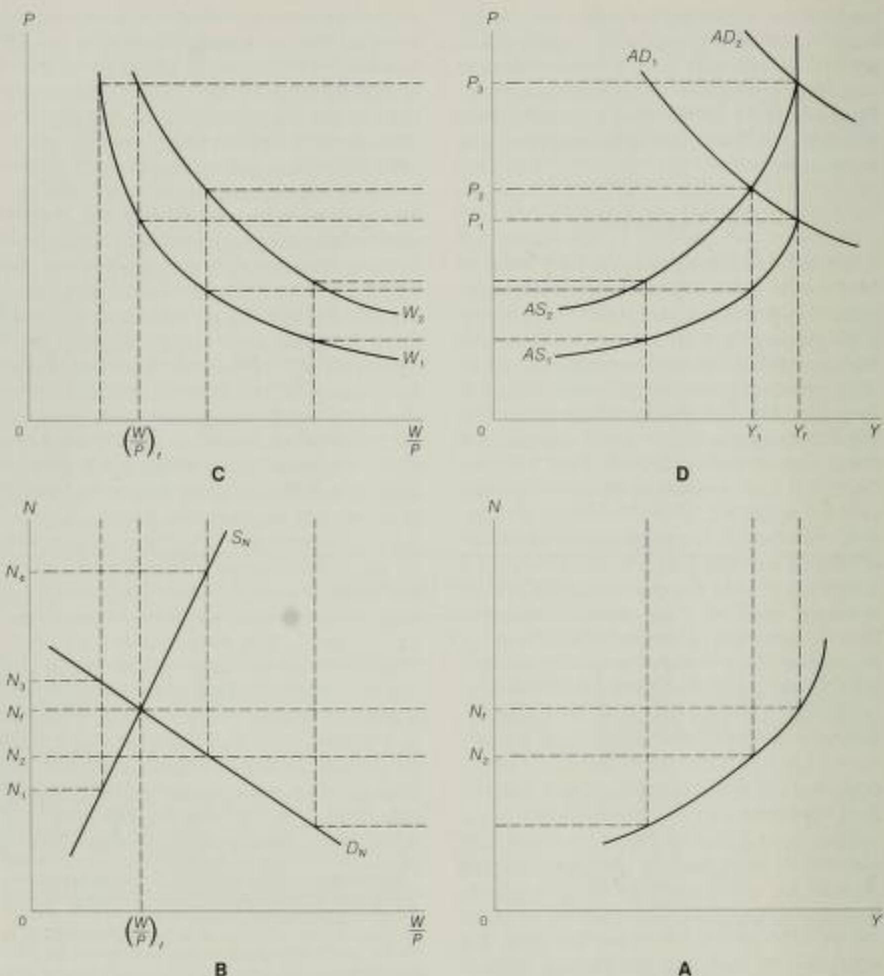


FIGURE 23-3
Cost-push inflation originating with autonomous money wage rate increase

in AS from AS_1 to AS_2 is a rise in the price level from P_1 to P_2 and a fall in the output level from Y_f to Y_1 as shown in Part D. In place of full employment of the labor force, N_f in Part B, there is now employment of N_2 and unemployment equal to $N_4 - N_2$. Thus, the rise in the price level

is accompanied by the appearance of unemployment.

Although not shown in Figure 23-3, one can see that further increases in the money wage will shift the *W* curve upward in Part C and, other things being equal, will bring about further upward shifts in the *AS* curve in Part D. With the *AD* curve given, once the adjustments have been worked out, each increase in the money-wage rate leads to a higher price level, lower output, and higher unemployment. If left to itself, such increases in the money-wage rate cannot continue indefinitely as the worsening unemployment that follows each such increase may be expected to restrain the unions' demands for ever higher money-wage rates. We will have more to say on this below, but it will be seen here that, at least over some range, a rising price level may result from rising money-wage rates.

Cost-push inflation is classified as wage-push if a rising price level is caused by rising money-wage rates. However, while a rising money-wage rate may be the cause of a rising price level, it does not follow that any rise in the money-wage rate will cause a rise in the price level. In other words, not all increases in money-wage rates are inflationary. Bypassing the complications that lie behind it, the basic relationship is that percentage increases in the money-wage rate that do not exceed the percentage increases in the marginal productivity of labor are not inflationary. Given the rule for firms' profit maximization under competitive conditions, $W = P \cdot \text{MPP}$, if W rises by 5 percent in any time period but MPP also rises by 5 percent, W/MPP (i.e., marginal cost) remains unchanged. Firms will offer any particular quantity of output at the same price level as before, or an unchanged P is consistent with profit maximization at any given level of output.

Without tracing the relationships in detail, what basically occurs in terms of Figure 23-3 is a rightward shift in the production function of Part A, the result of technological advance or an increase in the capital stock. This raises the MPP at each level of N and shifts the D_N curve in Part B to the right. With a given money-wage rate in Part C, the result will be a downward or rightward shift in the *AS* curve of Part D.[9] If an increase in *W* which *in itself* causes the *AS* curve to shift upward is just matched in percentage terms by an increase in the MPP which *in itself* causes the *AS* curve to shift downward, the two are offsetting and the price levels at which the various output levels are supplied by industry as a whole remain unchanged. Of course, to the extent that the percentage changes differ, the AS curve, on balance, will shift upward or downward, although recent experience has shown no cases in which *W* did not rise at least as fast as MPP and thus no cases in which the AS curve, on balance, shifted downward.

That increases in money-wage rates matched by increases in labor productivity are not inflationary is a fact and perhaps a familiar one, but is it also a fact that increases in money-wage rates unmatched by increases in labor productivity are necessarily inflationary? This clearly looks like all one needs to identify a wage-push, but that is not necessarily the case. The rise in the money-wage rate may not be the result of the exploitation of the power held by labor unions but the result of an excess demand for labor. In this event, the money-wage rate would rise whether there were labor unions or not and the rise cannot meaningfully be described as inflationary.

To see this, start in Part D of Figure 23-3 with an original equilibrium of Y_f and P_1 established by the intersection of AS_1 and AD_1. Assume then a shift in AD to AD_2, which is purely inflationary and pulls P up from P_1 to P_3. This, of course, is a standard case of demand-pull inflation covered in an earlier section. However, note the developments that follow from this. The higher P clearly means a larger profit per unit of output and a larger total profit for firms because the total receipts from the sale of an unchanged amount of output have risen pro-

[9] The process is illustrated by Figure 14-3, p. 240, although the model there assumes a perfectly inelastic *AS* curve instead of the one here which is less than perfectly inelastic over a range.

portionally with P. The greater profits spur firms to expand output, and they seek more labor for this purpose. With P at P_3 and W for the moment at W_1, W/P has declined to the extent that the amount of labor demanded increases from N_f to N_3 and the amount supplied decreases from N_f to N_1 in Part B. There is an excess demand for labor of $N_3 - N_1$. This excess demand forces W from W_1 to W_2 in Part C. With P at P_3, the increase in W from W_1 to W_2 raises W/P to its original level, $(W/P)_f$, and eliminates the excess demand for labor. At the same time, the increase in W shifts the AS curve upward from AS_1 to AS_2. At the new equilibrium, the supply of goods equals the demand for goods, and the supply of labor equals the demand for labor.

In this sequence, there has been a rise in the money-wage rate with no rise in the productivity of labor, and consequently there has been a rise in the labor cost per unit of output. However, notice that the rise in labor cost per unit of output is not the *cause* of the rise in the price level but the *result*. The money-wage rate rose as a result of an excess demand for labor derived from the upward shift in the aggregate-demand function for goods; the rise in the money-wage rate would have occurred whether labor was organized or not. Although this conclusion may be obvious, recognition of it is necessary if we are to see the fallacy in the argument of those who contend that any rise in wage rates in excess of the rise in labor productivity causes a rise in the price level.

The concept of wage-push inflation must then be limited to increases in labor costs that are the cause and not the result of higher prices. Wage-push inflation can follow only from "spontaneous" or "autonomous" upward shifts in the supply function, as opposed to those that are "induced" by excess demand for labor. Induced shifts may occur with or without a strong labor movement, but spontaneous shifts require organized labor with sufficient strength to push up wage rates in the absence of any excess demand for labor. Where unions have this strength, there is the possibility of a rise in wage rates that may produce a spontaneous shift in the supply function from AS_1 to AS_2, even though there is no shift in AD. As a first approximation, therefore, we can say that a rise in prices is the result of wage-push if the existing AD function is not able to support the rise in wage rates without a reduction in output and employment. This is illustrated by the reduction in output from Y_f to Y_1 that results from the shift from AS_1 to AS_2. On the other hand, a shift from AS_1 to AS_2 will not reduce output if that shift is induced by excess demand for labor that follows from a shift in AD from AD_1 to AD_2.

Wage-push inflation is impossible in an economy in which wage rates are determined by purely competitive market forces. In such a situation wage rates will rise or fall only in response to variations in labor supply and demand, the latter in turn depending on variations in aggregate demand for final output. However, it is not necessary that the labor force be completely unionized for the conditions needed for wage-push to arise. It is sufficient that a substantial portion of the labor force be organized and that the strength of these organizations be great enough to force increases in money wages that exceed productivity increases. With a labor force partially organized and partially unorganized, as is the case in the U.S. and most other free economies, and to the extent that nonunion wages are closely tied (with established differentials) to union wages, it is possible for wage-push to originate in unionized industries and for the higher wages gained there to spread to other industries.

There are a number of reasons to expect this spread of higher wages to nonunion workers. Employers of nonunion labor may raise the wages of their employees in order to discourage unionization, prevent employee discontent, and avoid loss of valued employees. However, to the extent that nonunion wages are *not* adjusted to union wages in these and other ways, a wage-push process in unionized industries leads to an ever greater gap between union and nonunion wages. This gap will mean relatively

higher prices on the goods produced by unionized industries, which may induce a shift of demand away from these goods in favor of the relatively lower-priced goods turned out by nonunionized industries. This gap also increases the possibility that new firms will enter the unionized industries and operate with nonunion labor. Developments such as these represent threats to the jobs and the higher wage rates of union employees and as such must act as a restraint on the push for higher wages in the unionized industries. Therefore, a wage-push inflationary process that creates a constantly widening gap between union and nonunion wage rates probably cannot be sustained. In sum, the initiation of an economy-wide wage-push inflation requires at least a partially organized labor force. The rate at which such an inflation will proceed and whether it can be sustained depends on, among other factors, the extent of unionization and the degree to which union-won gains spread to the balance of the labor force.

These results are based on the assumption that employers post higher prices for their products as rapidly as the forced increases in wage rates raise the cost of those products. Although the presence of labor unions on the supply side of the labor market is a prerequisite to wage-push inflation, the success of a union drive to force up wages depends in part on the demand conditions faced by the firms subjected to these wage demands. There is the individual firm, operating under monopolistically competitive conditions, that cannot raise prices to offset higher wage rates without losing much of its sales, for under monopolistically competitive conditions the demand curve for its output is quite elastic. Whether they are organized or not, workers cannot force such an individual firm to grant higher wages without eventually driving it out of business (unless the particular firm enjoys some special advantage that provides extraordinary profits). If, on the other hand, the same increase in wage rages is secured from all firms in a monopolistically competitive industry, wage-push is at least possible. Despite the elasticity of the demand curve facing each firm, no one firm need fear loss of sales to competitors by raising prices to cover higher wages, since all other firms in the industry may be expected to follow a similar policy. It still remains that the industry as a whole faces a less than perfectly inelastic demand curve. Therefore, if wage rates do not show similar rises in other industries, the industry in question may lose sales to others. Nonetheless, labor may still be able to force through a wage-push, albeit at the cost of some decrease in the output of this industry and in the employment it provides. Since the demand curves facing oligopolists are less elastic than those in monopolistically competitive markets, such firms present unions with the best opportunities for securing wage increases in excess of productivity increases with a minimum loss of jobs. Largely for this reason, wage-push, if it occurs, tends to be most pronounced in unionized oligopolistic industries. Although initiated in and offset by higher prices for the output of these industries, higher wages usually spread to other unionized industries as unions there seek to follow the "pattern" and to match the gains that have been won by other unions. The same increases may in turn spill over in large degree into nonunionized industries, as we mentioned earlier.

Thus, starting out with the existence of strong aggressive unions in the major oligopolistic industries, one can see how wage pressures originating there and wage gains won there can possibly spread through the system as a whole to produce some degree of wage-push inflation in the absence of any increase in current demand for output of the economy as a whole.

Profit-Push Inflation

Profit-push is another variant of supply inflation. Just as labor unions may exercise their market power by forcing wage increases, so oligopolists and monopolists may, in their drive toward greater profits, raise prices more than

enough to offset any cost increases. Again, just as labor unions are a prerequisite to a generalized wage-push inflation, so the existence of imperfectly competitive markets in the sale of goods and services is a prerequisite to profit-push inflation. Where prices of goods are set by the competition of buyers and sellers, as in agricultural commodities and raw materials, the seller cannot do very much about the price at which he sells. But many goods do not move through such markets, and the sellers of these goods "administer" prices. In an economy in which so-called administered prices abound, there is at least the possibility that these prices may be administered upward faster than costs in an attempt to earn greater profits. To the extent that such a process is widespread, profit-push inflation will result.

Although there is this similarity between the administered prices at which labor unions supply labor to firms and the administered prices at which firms supply goods and services to their customers, the responsibility for supply inflation has still generally been placed with labor unions. One reason for this lies in the differences between the wage-setting process followed by unions and the price-setting process followed by business. It is argued that unions commonly press for higher wages with an objective that is little more specific than Samuel Gompers' classic goal of "more." This is either because unions regard the demand curve for labor as highly inelastic or because they are more concerned with higher wage rates than with the amount of unemployment that higher wage rates will produce with given aggregate demand. In contrast to the officers of labor unions, those who administer prices of goods have a more or less definite objective in the form of profit maximization. Profits of a firm depend not only on prices but on sales and unit costs as well, and the latter depend in part on prices charged. Thus, although there is always the possibility of raising prices where prices are administered, the argument is that market realities enter more systematically into the setting of prices than into the setting of wage rates by labor unions. Firms with administered prices may generally respond promptly to wage increases by raising prices, but they are unlikely, when faced with unchanged demand and stable costs, to find cause to raise prices above those that are already "equilibrium" prices according to the profit maximization or other objectives of the firm.

Critics of business who assert that "giant corporations" price their output to squeeze out the last dollar of profit sometimes also accuse the same companies of pricing their output in a way that produces profit-push inflation. From the preceding, it should be clear that one can not have it both ways. If the companies have already fully exploited whatever market power they have, they can not also price their output in such a way as to produce a profit-push inflation without some sacrifice in those profits. On the other hand, if they have a reserve of unexploited market power, they can then price in a way to produce a profit-push inflation. But in this case the charge that they regularly squeeze out the last dollar of profit can not also be supported.

A qualification to this line of argument appears if the degree of market power held by business increases continuously over time or over certain intervals of time. There is no completely satisfactory measure of monopoly or oligopoly power, but the one most commonly used is the concentration ratio, the share of an industry's total output accounted for by its 4 largest firms. A study of 209 manufacturing industries shows that from 1947 (or 1954) to 1963 the concentration ratios in 85 increased, in 81 decreased, and in 43 remained relatively stable. Based on later data, from 1947 (or 1954) to 1967, for the same 209 industries the ratios increased in 95, decreased in 75, and remained relatively stable in 39.[10] Another study of 213 manufacturing industries shows that the average concentration ratio increased from 41.2

[10]See *Concentration Ratios in Manufacturing, 1967, Special Reports*, 1970, Part 1, Table 5, Bureau of the Census, Department of Commerce.

percent in 1947 to 41.9 percent in 1966, an increase of 0.7 percentage point. The 132 of these industries classified as capital goods industries showed a decline in the average ratio over these years from 45.1 to 43.4 or by 1.7 percentage points, and the 81 classified as consumer goods industries showed a rise from 34.8 to 39.6 percent or of 4.8 percentage points.[11] Evidence like this suggests some overall increase in concentration in manufacturing for this particular period of time but not an increase of major magnitude. Some of the U.S. economy's bouts with inflation over this period have been attributed to cost-push forces, but evidence like this does not lend much support to the argument that the cost-push was of the profit-push variety.

Cost-Push Inflation— The Problem of Control

Restrictive monetary and fiscal policies are the standard weapons used to curb demand-pull inflation. If a burst of investment spending, foreign purchases, or other such forces shifts the *AD* curve in Figure 23-1 upward from AD_1 to AD_2, the resulting rise in the price level may be countered by restrictive monetary and fiscal policies. In carrying out such policies, there are difficult questions of timing and degree, but there is little question that such policies are appropriate to combat demand-pull inflation.

In a supply inflation, however, restrictive monetary and fiscal policies are not so clearly appropriate. Such measures have their immediate impact on aggregate demand, but supply inflation is not the result of aggregate demand pressing against the economy's full-employment output. As a matter of fact, in attempting to distinguish between demand and supply inflation, the most telling evidence of a supply inflation is a rising price level with output appreciably below the level indicated by full employment. Under these conditions, reliance on restrictive monetary and fiscal policies may, by reducing aggregate demand, actually aggravate the inflationary pressures rather than eliminate them—for restrictive policy may restrain the rate of investment spending and thereby slow the rise in labor productivity that could otherwise offset, at least in part, a wage-push, which by definition is a more rapid rise in the money-wage rate than in labor productivity. On the other hand, a restrictive policy may succeed in checking a wage-push inflation if it reduces aggregate demand and output by an amount sufficient to create enough unemployment to prevent wage increases in excess of productivity increases. The problem here is that to avoid wage-push inflation by maintaining whatever unemployment rate is necessary for that purpose may require an unemployment rate that is socially and economically unacceptable. In other words, we might be purchasing price stability only at the cost of considerable social distress and a slowed rate of economic growth. If sustained periods of 7 or 8 percent unemployment are necessary to achieve price stability, it would seem to many people the lesser of two evils to accept the inflation that accompanies a 5- or even 4-percent unemployment rate.

The U.S. experience over the past decade and a half provides several illustrations of the problem of adjusting stabilization policies to conform with different types of inflation. The inflation during 1965–68 could be explained largely by the excessively rapid growth of aggregate demand. Of course, this growth was, in turn, initiated by the rapid growth of military expenditures that got under way with the beginning of the Vietnam buildup in 1965. Prices rose considerably over these years, but output also rose quite sharply; the rise in output was sufficient to push unemployment below 4 percent in 1966, the first year a rate below 4 percent had been achieved since the Korean War. However, with further reductions to a 3.6 unemployment rate by 1968, most of the slack in the economy had been absorbed and further in-

[11] See *Studies by the Staff of the Cabinet Committee on Price Stability*, Washington, D.C., January 1969, p. 58.

creases in aggregate demand could be expected to exert their impact predominantly on prices with little gain in output. If the existing inflation was not to go from bad to worse, it became imperative that demand be brought under control. Although the policy pursued by the government may surely be criticized for coming too late and perhaps also for being too little—action had been recommended by many economists as early as 1965—the increase in corporate and personal income tax rates through the 10-percent surtax imposed at long last by the Congress in 1968 was a step in the right direction. There could be little doubt that demand-pull forces were at work in the ongoing inflationary process, and the 1968 action was therefore an appropriate move.

In January 1969 the incoming Republican administration inherited the ongoing inflation, but it also inherited an economy with no serious unemployment problem. The January 1969 unemployment figure was 3.3 percent. The policy of the new administration, which was steadfastly adhered to for over two years, was to limit its attack on the inflation problem to what could be done with monetary and fiscal policies alone. Wage and price controls or guideposts or any other version of an incomes policy, even something as moderate as "jawboning," were rejected out of hand. The administration's opposition to such measures was almost theological in its fervor. Although the administration never spelled out exactly how this was to be accomplished, it held that through monetary and fiscal policies the rate of inflation would be reduced to an acceptable level without at the same time producing any appreciable rise in the unemployment rate. What actually happened is, of course, now in the record books: The economy went through its first recession since 1960–61 and suffered a rise in the unemployment percentage from an average of 3.5 in 1969 to 4.9 in 1970 and 5.9 in 1971. In exchange for this sacrifice, the economy received no substantial slowing in the rate of inflation. The restrictive monetary and fiscal policies turned out to be not without effect, but the effect was predominantly on output and employment and only moderately on prices, just the opposite of what was to have been expected according to the administration's so-called "game plan."

The difficulty arose from the fact that what by 1969 had come to be an inflation driven in large part by wage-push forces was being attacked by the administration with weapons that are not well suited to combat this kind of inflation. As we have seen, the wage-push element in an inflationary process can only be checked by restrictive monetary and fiscal policies through the capacity of those policies to restrain output and raise unemployment enough to prevent wage increases in excess of productivity increases. It is this fact that gave rise to the widespread skepticism among economists that the administration could, by resort to any conceivable combination or variety of restrictive monetary and fiscal policies, check the existing strong inflationary movement without producing severe unemployment. Beyond this, there was considerable doubt whether even a pronounced increase in unemployment would be adequate to the task. Given the fact that the inflation had become progressively worse since 1965 and the fact that the public quite generally expected more of the same, labor unions were understandably determined to settle for nothing less than wage rate increases that not only covered productivity gains but also provided whatever was needed to catch up with past price increases and to cover the prospective price increases for the period of the contract. Wage rate increases actually continued far in excess of productivity increases in 1970 and the first seven months of 1971 that preceded the introduction of wage and price controls, despite the fact that the unemployment percentage rose from 3.5 in 1969 to 4.9 in 1970 and to 6.2 percent for the first seven months of 1971. The prerequisite for eliminating whatever influence wage-push was exerting on the price level, namely, to restrain the rate of wage increase to the rate of productivity increase,

was not met, even though the unemployment percentage had been pushed up drastically in the attempt.

Whether a continuation of the same policies for another few months would have begun to produce the intended results will never be known, although many within the administration believed it would have. In any event, the adoption by the administration in August 1971 of an incomes policy or a policy of direct intervention in wage and price setting amounted to an admission that the monetary and fiscal policies that had been relied on over the preceding two and a half years had not worked as expected. Without entering into the question of the degree to which the wage and price controls, maintained in various forms until April 1974, were in their turn successful, what is relevant here is that a policy of controls at least offered some hope of slowing down a wage price spiral without putting the economy "through the wringer," something that restrictive monetary and fiscal policies are unable to do where strong cost-push forces are at work.

From August 1971, the month in which controls were introduced, to December 1972, the Consumer Price Index rose at an annual rate of only 3.3 percent, but from December 1972 to December 1973 the rate jumped to 8.8 percent and from December 1973 to December 1974 reached the "double digit" rate of 12.2 percent. For the following year the rate declined to 7.0 percent and for the year ending December 1976 to 4.8 percent. Like the sharp rise in prices in 1969–70, the more recent surge of prices, especially the "double digit" experience of 1974, is another episode of cost-push inflation. However, while wage-push was to some degree exerting its influence during these years, the cost-push forces at work during 1973–74 were in large part of a special, nonrecurrent nature. As 1974 was a year of worldwide economic recession, 1973 was a year of world-wide economic boom. At the same time that they enjoyed high prosperity at home, most of the industrialized countries suffered poor agricultural crops during 1973, which led to a great surge in demand for U.S. farm output. The farm products component of the U.S. wholesale price index shot up 36 percent during 1973! The resulting rise in the overall U.S. price level was only one part of the inflation of that year accounted for by international transactions. Another part resulted from the two devaluations of the U.S. dollar in December 1971 and February 1973. Import prices, which had risen at a 7-percent rate from 1971 to 1972 more than doubled to about 17 percent from 1972 to 1973. While the devaluations had the intended effect of stimulating U.S. exports, they were responsible for a large part of the increase in import prices and through this contributed significantly to the inflation during 1973.

A special factor without parallel came into the picture in 1974: the quadrupling of oil prices by the OPEC cartel following the lifting of the embargo on oil shipments to the U.S. by Arab oil-producing nations imposed in October 1973. The 44-percent rise in U.S. import prices in 1974 made what had been a very sharp rise of 17 percent in the preceding year look small in comparison. As the huge increase in the price of imported oil worked its way through the economy, not merely in the form of higher gasoline prices and higher home heating fuel prices but in higher prices of the hundreds of other products for which oil is a basic raw material, its impact on the general price level in 1974 was of major importance. Estimates of the impact differ, but it is safe to say that the "double digit" inflation rate of 1974 would not have occurred in the absence of this explosive increase in costs imposed by OPEC on the U.S. economy as well as on those of other industrialized oil-importing countries. Although this is another matter, there are those who maintain that it was the sudden severe jolt of the huge increase in payments made for oil that was primarily responsible for the worst recession that the U.S., western Europe, and Japan have suffered since the nineteen-thirties.

In terms of their impact on the domestic econ-

omy's price level, special factors like those noted here operate in the same way as other cost-push factors like wage-push. The higher costs bring about an upward shift in the aggregate supply function of the kind illustrated in Part D of Figure 23-3 by the shift from AS_1 to AS_2. Higher prices for OPEC petroleum raise the price level at which each level of output will be supplied as surely as do higher money-wage rates unmatched by higher labor productivity. However, while a policy of wage and price controls may be able to restrain, at least for some period of time, cost-push of the wage-push or profit-push varieties, it is unable to do much at all when confronted with cost-push due to special factors like those considered. The U.S. government obviously can not impose price controls on the goods supplied by foreign producers. It can do something to control increases in prices due to sudden, large increases in foreign demand, e.g., grain prices in 1973, by imposing an embargo or setting limits on quantities that may be exported, but in the case of agricultural products this runs into the political problem of enraging the nation's farmers.

The special factors noted here are primarily one-time factors which lifted the price level to a higher plateau, a much higher plateau, over the period of a couple of years. The OPEC cartel, if it survives, will obviously continue to raise prices, but there will just as obviously not be another quadrupling or even a doubling in one year. Unless other special factors now unforeseen appear, e.g., formation of OPEC-type cartels for other vital imported raw materials, the problem of cost-push inflation will again be essentially the problem of wage-push. And the question of control will again be whether or not to resort to restrictive monetary and fiscal policies which worsen the decline in output and employment that automatically come with wage-push in the hope that the greater decline will quickly put an end to the wage push. Because there is no great likelihood that a policy of deliberately making a bad situation temporarily worse will in short order make it better, restrictive monetary and fiscal policies will not likely be the policies adopted to combat wage-push inflation. Allowing for all of its costly side effects, the policy of wage and price controls may on balance be no better, but it is as a matter of practice the major alternative to which governments turn when they decide to pursue an active anti-inflationary policy.

This does not deny that there are any number of policies of narrower scope whose implementation can contribute to some degree to a slowing of inflation. These include actions to improve the functioning of labor markets through retraining programs, improved dissemination of information on job vacancies, and relocation grants to the unemployed. Also contributing an anti-inflationary increase in labor productivity would be the reduction of "featherbedding" in union contracts, although this is most difficult to achieve. Among other measures are those which strengthen competition in goods markets. Government can move toward correcting policies that encourage inefficient business operations with their resultant high costs of production. For example, regulation of the interstate trucking industry has had the effect of stifling competition, protecting the inefficient, and creating great waste, the cost of which is borne by the final consumer. In the area of foreign trade, movement toward freer trade rather than protection can be undertaken with the benefit of lower prices to the consumer. To the degree that profits of firms and the jobs of their workers are to be protected by government from foreign competition, this should be done directly instead of through higher prices paid by consumers across the country. Measures like these will not eliminate inflation, but they will make a contribution. Furthermore, they are measures which commend themselves apart from whatever help they may provide in controlling inflation. However, despite the less than successful experience with wage and price controls in the past, the fact is that it is to some such system of controls that governments turn as the alternative to monetary and fiscal policies when these policies are unsuited to combating the kind of inflation confronted.

COST-PUSH INFLATION: ITS RELATION TO DEMAND-PULL

Up to this point we have examined separately the two parts of a well-established dichotomy in inflation theory: demand-pull and cost-push. Although this dichotomy is now a part of the language of economics, some economists object to its implication that an inflation is *either* demand-pull *or* cost-push. They argue that any actual inflationary process contains some elements of both. Expressed in this fashion their argument can hardly be denied. However, if the dichotomy is accepted as nothing more than a convenient twofold classification of types of causation, their objections do not apply: It is at least helpful in separating two distinct sets of forces that are usually simultaneously and interdependently at work in any actual inflationary process.

In terms of this dichotomy, it should be noted that there is a lack of symmetry between the demand-pull and cost-push theories. An inflationary process may begin with generalized excess demand and may be expected to persist as long as excess demand is present, even though no cost-push forces whatsoever are at work. Excess demand will raise prices, which in turn will raise wage rates, but the rise in wage rates in this case is not the result of cost-push. Notice, however, that this does not rule out the possibility that cost-push forces may also be at work to produce an even greater rise in wage rates. On the other hand, an inflationary process may begin on the supply side, but it will not long persist unless there is an increase in demand. For example, an autonomous rise in wage rates will raise prices in the absence of any increase in demand. For a cost-push inflation so initiated to be sustained, however, one wage increase must be piled on top of another; but, in the absence of an increase in demand this would mean ever smaller production and ever greater unemployment. Sooner or later this must limit any inflationary process that depends on changes on the supply side alone.

This asymmetry can be illustrated by Figure 23-4. With output at Y_f, shifts in the aggregate demand function from AD_1 to AD_2 to AD_3 and beyond can carry the price level ever higher, from A to E to G, and so forth, in a sustained inflationary process. With full employment, wage rates will rise along with prices as producers, encouraged to expand output by the enlarged profits that result from the rising aggregate demand, increase their demand for labor. As long as the forces feeding the demand for final output continue to shift the AD function ever higher, inflation will continue unchecked. In the extreme case, a runaway price level known as "hyperinflation" may result. However, starting again from Y_f, a wage-push or a profit-push

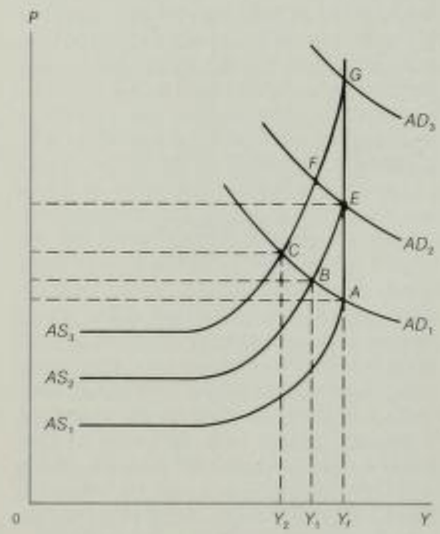


FIGURE 23-4
Shifts in the aggregate supply function and cost-push inflation

that shifts the aggregate supply function from AS_1 to AS_2 will, with the AD function still at AD_1, produce an intersection at B and reduce output below the full-employment level. A further upward push on the supply side to AS_3, unless accompanied by a shift in AD above AD_1, will move the intersection to C and further reduce output and employment. The successive reductions in output and the growing unemployment that result under these conditions will bring the inflation to an end. Thus, unlike demand-pull, inflation may originate on the supply side, but it cannot be sustained unless there is an appropriate increase on the demand side.

The crucial question, of course, is what shift, if any, occurs in the aggregate demand function as the aggregate supply function shifts in the way described. At one extreme we may find the AD function shifting upward proportionally with the upward shift in the AS function. If we take rising money wages as the cause of the upward shift on the supply side, such a result assumes that as firms raise prices in response to higher wage rates, aggregate spending rises proportionally to maintain output and employment unchanged at the higher price level. This is equivalent to no decrease in real demand, a result that can follow only if there is, for example, a sufficient expansion of the money supply and so no rise in interest rates that would reduce real investment spending. Or, combining the interest-rate effect and other effects like the Pigou effect, this result can follow only if these effects do not produce a decrease in real demand or, to phrase it differently, if they do not prevent a rise in aggregate spending proportional with the rise in the price and wage structure. But, because of these effects and for other reasons, it seems most likely that aggregate spending will *not* rise in line with the rise in wages and that as a result there will be some decline in output and employment and a lesser rise in prices than would be the case on the assumptions above. To the extent that this is the economy's characteristic response to the cost-push, successive cost-pushes can indeed force up wages and prices, but only at the cost of reduced employment and output. Thus we reach the conclusion noted above: Pure cost-push cannot in itself produce a sustained inflationary process.

If the conclusion just reached is correct, the prospects of an inflationary process being fed in a sustained way from the supply side alone seem quite limited. But this overlooks one crucial consideration: Although a cost-push will probably not lead to a spontaneous increase in aggregate demand of the extent needed to prevent a reduction in output and employment, the required rise in aggregate demand may follow as the result of expansionary monetary and fiscal policies. Because cost-push left to itself can produce a reduction in output and aggravate any existing unemployment, it is likely that this very fact may lead the monetary and fiscal authorities to pursue policies designed to support output and employment even though the result of such policies is an inflationary process that cost-push forces alone could not bring about. This brings us face to face with a well-known dilemma: the apparent impossibility of simultaneously achieving full employment and avoiding some amount of inflation in an economy in which strong wage-push and/or profit-push forces are at work.

Cost-push, viewed from this perspective, can thus mean a sustained inflationary process to the extent that "full employment at all costs" becomes the overriding economic objective of public policy. Such a policy requires the "validation" of an upward push of wages in order to prevent what otherwise would be an unacceptably high level of unemployment, but the validation of one wage-push is an open invitation to labor unions to bring off another wage-push. In terms of Figure 23-4, a "ratchet" is traced out by the movements from A to B to E to F to G as the price level moves up in a persistent inflationary process. This same process, it should be noted, can begin and proceed in an economy below full employment. Starting at B, a wage-push might shift us to C. An upward shift in AD

from AD_1 to AD_2 moves us to F, at which another wage-push (not shown) could shift the AS function upward another notch. As long as the policy is to validate each wage-push in the effort to prevent the worsening of an already existing unemployment problem, a sustained rise in the price level may occur in an economy operating below full employment. Because the choice typically made by government is to validate the cost-push, the existence of labor unions and oligopolistic business with the market power to force up the price level and the willingness to exploit that power produces a combination that can sustain an inflationary process at the full employment position or at a position below full employment.

THE RELATIONSHIP BETWEEN THE INFLATION RATE AND THE UNEMPLOYMENT RATE

Since wage costs represent the backbone of the price structure, the study of supply inflation naturally leads one from the relationship between the rate of price increase and the rate of wage increase to the relationship between the rate of wage increase and the rate of unemployment of the labor force. If in addition to the direct relationship between the rate of price increase and the rate of wage increase, there is also an inverse relationship between the rate of wage increase and the rate of unemployment, it follows that there is an inverse relationship between the rate of inflation and the unemployment rate. Because the data for various countries and for various time periods do show this inverse relationship between the rate of wage increase and the unemployment rate, the unemployment rate in this indirect way becomes a factor of major interest in the study of supply inflation.

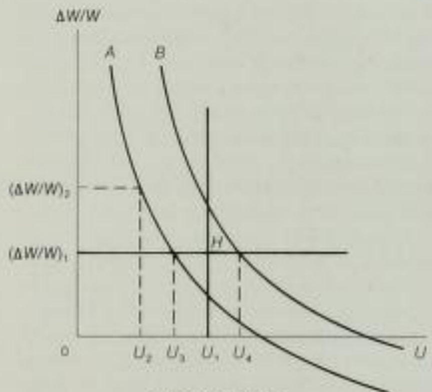


FIGURE 23-5
Phillips curves

The analysis of the particular relationship between the rate of wage increase and the unemployment rate proceeds in terms of the "Phillips curve," so named after A. W. Phillips, who pioneered in the investigation of this relationship for the United Kingdom.[12] In its simplest form, a Phillips curve may be derived from an economy's data for a period of years by plotting for each year the percentage of money wage rate increase, $\Delta W/W$ (vertical axis), against the percentage of the labor force that is unemployed, U (horizontal axis). A curve fitted to the points so plotted, for certain periods of years, e.g., 1961–68, will slope downward to the right like the hypothetical curves A and B in Figure 23-5.[13] Any curve of this general shape sug-

[12] See "The Relation between Unemployment and the Rate of Change in Money Wage Rates in the United Kingdom, 1862–1957" in *Economica*, Nov. 1958, pp. 283–99, reprinted in M. G. Mueller, ed., *Readings in Macroeconomics*, 2nd ed., Holt, Rinehart and Winston, 1971, pp. 245–56.

[13] We will look at some actual data and a curve fitted to them below.

gests that the rate of money wage increase is inversely related to the unemployment rate. If the actual curve happened to be A in Figure 23-5, a decline in the unemployment rate from U_3 to U_2 would raise the percentage wage increase from $(\Delta W/W)_1$ to $(\Delta W/W)_2$.

Explanations of the Phillips Curve Relationship

There are two principal explanations for this relationship. One is found in the behavior of organized labor. The argument here takes off from the one presented above under the heading of wage-push inflation. We saw there that organized labor has the power to bring about autonomous increases in wage rates in excess of increases in productivity. This leads to rising prices of goods in a process that is accordingly called wage-push inflation. To the argument that organized labor is able to push through autonomous increases in wage rates greater than productivity increases, we now add the following: The degree to which it is able to do this will vary inversely with the unemployment percentage and the ease of labor markets. Lower unemployment and tighter labor markets are conditions under which organized labor will become more aggressive and press for larger wage increases; the opposite conditions are those under which organized labor will be less demanding. Furthermore, because times of low unemployment and tight labor markets are ordinarily times of buoyant demand for goods and abundant profits, business will usually choose to grant the "excessive" wage increase demands rather than face the possibility of a strike and a shutdown of such profitable production. Under the opposite set of conditions—high unemployment and low profits—business would have much less to lose and would show considerable resistance to even moderate wage increase demands. With the relative bargaining strength of labor unions and employers varying in this way, what is to be expected is an inverse relationship between the percentage wage increase and the unemployment percentage of the kind shown by the Phillips curve.

A second explanation, which is general and not dependent on the relative bargaining powers of organized labor and business, runs in terms of excess demand for labor.[14] The explanation is somewhat involved and we will sketch only the bare essentials here. For this purpose, an abbreviated two-part graphic system is used instead of the four-part system needed to show the complete analysis graphically.

The explanation begins with supply and demand curves for labor as shown in Part A of Figure 23-6. For the curves as given, it is seen that the wage rate W_3 equates the supply of and demand for labor. Although this wage rate indicates an equilibrium in the labor market, it does not indicate an absence of unemployment. It is an equilibrium in the sense that the number of unemployed workers is just equal to the number of vacancies that employers seek to fill. It would be an equilibrium with zero unemployment only if frictional unemployment were always zero. This would require, for example, that no worker ever changed jobs or, if he did, that not a day's work was lost in the process, a situation obviously not realized in practice.

At any wage rate below this equilibrium, e.g., W_2, the number of vacancies exceeds the number seeking jobs, or there is excess demand for labor, in this case equal to GH. In the same way, at any wage rate above the equilibrium, there are fewer vacancies than the number seeking jobs, or there is excess supply of labor. At W_4 excess supply is equal to MN. Looking specifically at cases of excess demand, a key argument in the present explanation is that the rate at which wage rates rise varies directly with the

[14] This explanation is the one advanced by R. G. Lipsey in "The Relation between Unemployment and the Rate of Change of Money Wage Rates: A Further Analysis," in *Economica*, Feb. 1960, pp. 1–31, reprinted in R. A. Gordon and L. R. Klein, eds., *Readings in Business Cycles*, Irwin, 1965, pp. 456–87.

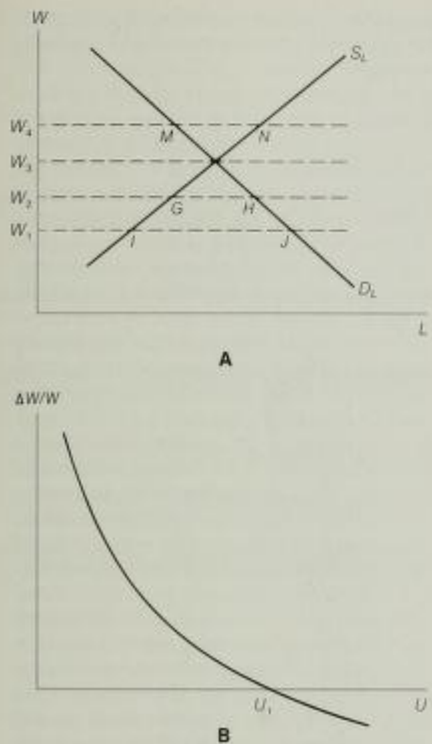


FIGURE 23-6
The labor market and the Phillips curve

extent of the excess demand for labor. Thus, with a wage rate of W_1 and an excess demand of *IJ*, there will be a more rapid rise in wage rates than with a wage rate of W_2 and the smaller excess demand of *GH*.

Another key argument is that the period of time required, on the average, for unemployed persons to find jobs varies inversely with the extent of excess demand for labor. In Figure 23-6, the average search time involved for workers without jobs to find the kind of job they are looking for will be shorter when excess demand is *IJ* than it will be when excess demand is *GH*. From this follows the important conclusion that the amount of unemployment will vary inversely with the amount of excess demand. In other words, with a given labor supply, whatever rate of unemployment exists at the equilibrium wage rate, that unemployment rate will be smaller as we move to successively lower wage rates or successively larger levels of excess demand. However, at the same time that the unemployment rate will be reduced by a greater excess demand for labor, a greater excess demand for labor will raise the rate at which wage rates increase. This brings in the earlier argument that how fast wage rates will rise in the face of excess demand depends on the magnitude of that excess demand. Putting together this relationship in which the rate of increase in the wage rate depends directly on the amount of excess demand and the rate of unemployment depends inversely on the amount of excess demand, we have the Phillips curve relationship, which shows the rate of wage increase and the unemployment rate to be inversely related.

To illustrate, suppose that when there is neither excess supply nor excess demand, i.e., a wage rate of W_3 in Part A of Figure 23-6, there is an amount of unemployment that gives us the unemployment rate U_1 in Part B of that figure. This unemployment rate is accompanied by a zero percentage rate of wage increase, as there is neither an excess supply of nor an excess demand for labor at the wage rate that gives us the unemployment rate of U_1. At any wage rate below W_3 we have excess demand. This means a lower unemployment rate or a leftward movement in terms of the horizontal axis in Part B, and also a higher percentage rate of wage increase or an upward movement in terms of the vertical axis in Part B. We thus find a movement to a point back up the Phillips curve. The greater the excess demand, the further from the originally selected point at U_1 will we be. The same reasoning may be applied to

a wage rate above W_3. In this case, there is excess supply, which means a higher unemployment rate than U_1 and a lower rate of wage increase than zero, or a movement to a point down the Phillips curve from the originally selected point at U_1.

The two explanations of the Phillips curve relationship noted here are not alternatives. The factors involved in the union power explanation and those in the excess demand explanation may be operating at the same time. The tendency of the rate of increase in money-wage rates to rise with a falling unemployment percentage may thus be explained in terms of both these causes as well as in terms of some lesser causes we have not entered into here.

Wage Increases and Inflation

To the extent that the actual data on the rate of wage increase and the rate of unemployment trace out a Phillips curve of the kind shown in the figures above, they add significantly to our understanding of inflation. A stable curve of this kind, if it exists, tells us what percentage rate of wage increase is to be expected with each rate of unemployment, and, with wage costs the backbone of the price structure, it indirectly tells us a good deal about what rate of inflation is to be expected with differing percentage rates of wage increase.

Not every rate of money wage increase is, of course, inflationary. Looking back at Figure 23-5, the horizontal line at the level $(\Delta W/W)_1$ indicates the percentage increase in labor productivity and therefore the percentage increase in the money-wage rate that, on the average, employment, a figure that most U.S. economists would perhaps place between 4 and 5 percent nowadays.

The point of intersection between the horizontal at $(\Delta W/W)_1$ and the vertical at U_1 is particularly significant. If the actual position of the Phillips curve is such that it intersects the horizontal line to the left of H, as the A curve does, it suggests that supply inflation of the wage-push type may be avoided through monetary and fiscal policy. Although the preceding section emphasized the inherent limitations of such a policy's ability to cope with wage-push inflation, an illustration of which was the 1969–70 experience, in the event that the Phillips curve is in the A position, monetary and fiscal policy may be employed without inflicting an unacceptably high unemployment rate on the economy. Thus, if the employment rate is as low as U_2, the wage increase rate is $(\Delta W/W)_2$, a rate in excess of the rise in productivity, which is equal to $(\Delta W/W)_1$, and so a rate of wage increase that indicates wage-push inflation. Resorting to restrictive monetary and fiscal policy in this case can end the wage-push inflation by reducing aggregate demand to a level at which the resulting unemployment rate rises to U_3. Since U_3 is still below U_1, it is still within the range of acceptability. Suppose, however, that the Phillips curve is one like B, which cuts the horizontal line to the right of H. Only by forcing on the economy the unacceptably high unemployment rate of U_4 will wage demands be held in line with productivity and wage-push pressures against the price level be held in check. In such a case, restrictive monetary and fiscal policy does not pro

of *H*. Studies for the U.S. economy during the fifties and early sixties indicated Phillips curves that are more like *B* of Figure 23-5, and accordingly they led to what in those years were regarded as pessimistic conclusions. One suggested that a 5- to 6-percent rate would probably be required to hold wage increases to no more than the productivity growth of the U.S. economy. Furthermore, to satisfy the requirement of no more than 3-percent unemployment (a commonly accepted definition of "full employment" during the sixties), prices might have to rise by 4 to 5 percent per year, figures that spell something more than creeping inflation.[15] Readings of the Phillips curve in the late sixties were still more pessimistic. They suggested a higher Phillips curve for which a rate of price increase of 5 percent could be expected to accompany a 4-percent unemployment rate rather than the 3-percent unemployment rate that had earlier been matched with this rate of inflation. These readings also suggested that to hold the rate of inflation down to a low 2 percent would require an unemployment rate of $5\frac{1}{2}$ percent.[16]

Through these years most economists accepted the idea of a stable tradeoff between the unemployment rate and the inflation rate: Society could exchange a higher inflation rate for a lower unemployment rate or a lower inflation rate for a higher unemployment rate on reasonably well defined terms. A look at the data shows part of the basis for such a conclusion. Figure 23-7 plots the rate of change of the average hourly earnings rate in manufacturing against the overall unemployment rate for the years from 1954 to 1976. Figure 23-8 does the same for the rate of change of the consumer price index against the overall unemployment rate. The full scatter of dots in these figures shows anything but a clear inverse relationship between wage rate change and the unemployment rate in the one case and the inflation rate and the unemployment rate in the other case. A clear relationship, of course, would be one in which all of the dots fall on or close to a smooth curve sloping downward to the right that can be drawn through them. However, while this is far from the whole story, what otherwise looks like an entirely random scatter takes on a different appearance when one allows for the role of the business cycle.

For example, referring back to the chronology of classical cycles in Table 19-1 (p. 363), an expansion began in August 1954, reached a peak in July 1957, and the following recession bottomed out in April 1958. Figure 23-9 shows separately the dots for the years of this cycle and also for 1959, the first year of the next expansion. As the unemployment rate fell during expansion from 1954 to 1957, the inflation rate tended to rise.[17] Then during the contraction, 1957–58, the inflation rate tended to fall. However, it did not tend to retrace the path it followed during the expansion. The post-World War II record exhibits an upward trend in the rate of inflation. Thus, in the present illustration, for any relevant rate of unemployment, the inflation rate is higher in the 1957–58 contraction period than it was at that rate of unemployment earlier during the 1954–57 expansion period. Some other cycles during this period show, at least to some degree, a similar kind of clockwise loop.

In the classical business cycle chronology, the expansion from February 1961 to November 1969 is the longest on record and it is the

[15]See P. A. Samuelson and R. M. Solow, "Analytical Aspects of Anti-Inflation Policy," in *American Economic Review*, May 1960, p. 192. Another study of this sort drew more pessimistic conclusions for the U.S. economy: See R. J. Bhatia, "Unemployment and the Rate of Change in Money Earnings in the United States, 1900–1958," in *Economica*, Aug. 1961, pp. 285–96.

[16]G. L. Perry, "Changing Labor Markets and Inflation," in *Brookings Papers on Economic Activity*, 3, 1970, pp. 411–41 and R. J. Gordon, "Inflation in Recession and Recovery," *ibid.*, 1, 1971, pp. 105–58.

[17]As measured by the consumer price index, the inflation rate not only did not rise from 1954 to 1955 but actually turned negative, the only year since 1954 in which this happened. If we were to work with the GNP deflator, we would find a slight positive change in the inflation rate from 1954 to 1955 instead of the negative change shown by the consumer price index.

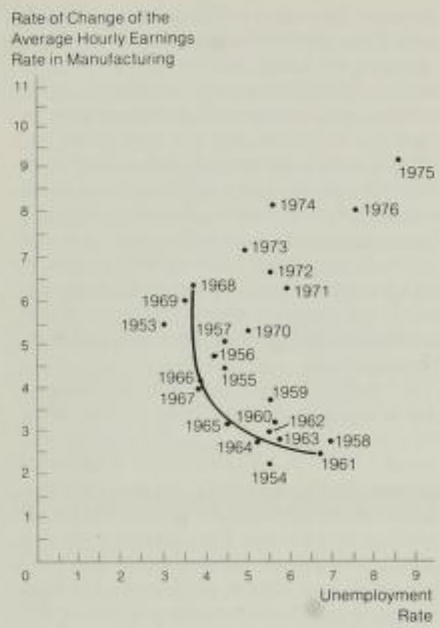


FIGURE 23-7
Rates of change of wage rates in manufacturing and the unemployment rate, 1954–1976

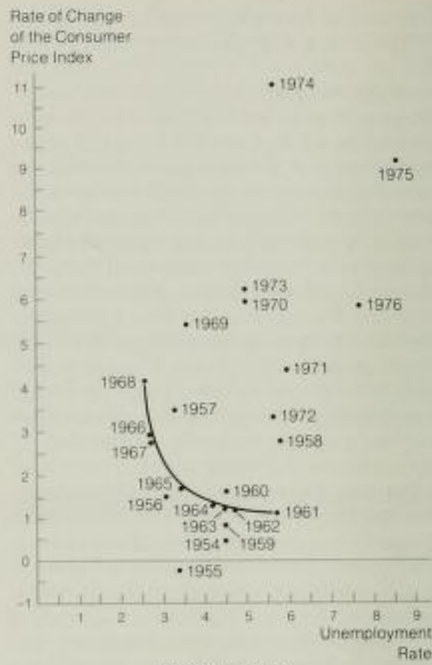


FIGURE 23-8
Rates of change of the consumer price index and the unemployment rate, 1954–1976

years of this expansion from 1961 to 1968 which more clearly than any others reveal a tight inverse relationship between the rate of wage increase and the unemployment rate and between the inflation rate and the unemployment rate. In Figures 23-7 and 23-8 a curve has been fitted to the dots for these years.[18] It was the occurrence of so systematic a relationship over these particular years that lent strong support to the conclusion of a definite trade-off between the two variables.

However, once one goes to 1969 and on into the seventies, the picture is drastically different. Figure 23-8 does not show for these later years anything resembling the close relationship found between the inflation rate and the unemployment rate for the 1961–68 period. One might argue that the dots in Figure 23-8 show, on the average, a very loose inverse relationship for the business cycle expansion period 1970–73, but the relationship for these years may be so weak as to be meaningless. How-

[18]While the curves in the two diagrams for the 1961–68 periods are very similar in shape, it will be noted that the rate of price level increase corresponding to any of the unemployment rates below the curve in Figure 23-8 is less than the rate of wage rate increase for the same unemployment rate in Figure 23-7. This, of course, reflects the point made above: Not all money wage rate increases are inflationary. The difference between the heights of the two curves basically reflects the average annual increase in labor productivity.

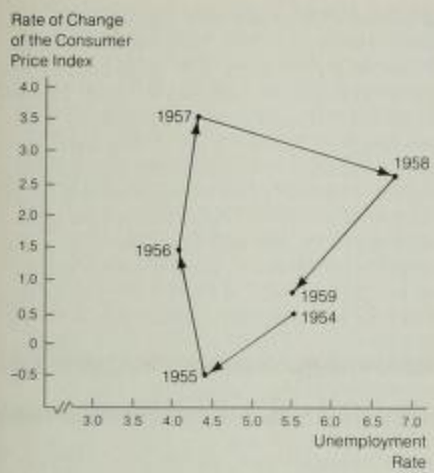


FIGURE 23-9
Rate of change of the consumer price index and the unemployment rate, 1954–1959

ever, if it is at all meaningful, the Phillips curve that emerges is one that has shifted upward substantially from the position it occupied from 1961 to 1968. In other words, although we did not return in these years of expansion to anything like the inflation rates found for each unemployment rate during the 1961–68 expansion, there is some evidence during these years that the inflation rate varied inversely with the unemployment rate, but a higher rate of inflation than before was now associated with each unemployment rate.[19] Those who look at the figures in this way find in them some support for the tradeoff argument. On the other hand, others conclude from the scatter of dots since 1969 that the conventional Phillips curve relationship has broken down completely. Some of the last few years, especially 1974–75, were extraordinary years during which the economy suffered its worst recession since the thirties at the same time that it absorbed great upward price pressures from external sources. Perhaps the late seventies and eighties will be years during which the record will show a reemergence of a distinct inverse relationship between the inflation rate and the unemployment rate or perhaps it will show that the Phillips curve is indeed no more. This is an empirical question and only the record of the years ahead can provide an answer.

The "Non-Trade-Off" Phillips Curve

Although the data for the sixties suggest a systematic trade-off between the inflation rate and the unemployment rate as we have seen from the curve fitted to these years in Figure 23-8, there were some economists who in the sixties disputed the conclusion that so many other economists had reached on the basis of this information, namely, that this curve was one along which policymakers could choose one position or another as they saw fit. For until

[19]A number of economists in the early seventies noted an upward shift in the Phillips curve and identified several factors to explain this phenomenon. One is the shifting composition of the labor force—notably an increase in the proportion of teenagers and women—and in the unemployment experience of differing age-sex groups. Because teenagers and women have a higher unemployment experience, the increase in the fraction of the labor force made up of these groups means that any given unemployment rate is associated with a tighter overall labor market than would otherwise be the case. With the rate of change in wages reflecting the tightness of the labor market, the result is that for any unemployment percentage this rate has been higher in recent years than it was a decade or more ago. This in turn leads to the finding noted earlier that the rate of inflation associated with any given rate of unemployment has been higher in recent years than earlier. Another factor is the increasing sensitivity of wage demands to inflation. As noted in a different connection in the preceding chapter, the experience of labor with inflation from 1965 onwards had by 1970 or sooner led labor to take inflation into account in bargaining over money-wage rates. At any given unemployment rate, wage demands are greater than they would be in a less inflationary climate. This operates to shift the Phillips curve upward. See G. L. Perry, "Changing Labor Markets and Inflation," in *Brookings Papers on Economic Activity*, 3, 1970, pp. 411–41, C. L. Schultze, "Has the Phillips Curve Shifted? Some Additional Evidence," *ibid.*, 2, 1971, pp. 452–67, and R. J. Gordon, "Wage-Price Controls and the Shifting Phillips Curve," *ibid.*, 2, 1972, pp. 385–430.

the experience of the seventies was at hand, there was a generally held belief that the Phillips curve was a fairly stable relationship which offered a "menu of choices." For example, if the objective were to move from an existing 6-percent unemployment rate to a 4-percent rate, the goal could presumably be achieved if society was willing to trade the higher rate of inflation indicated by the Phillips curve at the 4-percent unemployment rate for the lower rate of inflation that is found at the 6-percent unemployment figure. In other words, according to these economists, if society was willing to pay the price in terms of inflation, public policy could be adopted to move the economy to a 4-percent unemployment rate or whatever other realistic rate might be selected as a goal.

But even before the belief in such a stable curve was dispelled by the experience of the seventies, these other economists had argued that the alleged "menu of choices" did not exist. They maintained that society could at best only *temporarily* trade off more inflation for less unemployment by pursuing policies to change the economy's position along the curve.[20] To them the downward-sloping curve relating the inflation rate and the unemployment rate is a short-run or transitory phenomenon that is valid only as long as discrepancies between anticipated and actual price changes prevail. According to these economists, once such discrepancies have been removed, the downward-sloping Phillips curve ceases to exist.

To illustrate their view, assume an existing 5-percent unemployment rate, U, and a 3-percent inflation rate, $\Delta P/P$, as shown by the point A in Figure 23-10. Government pursues expansionary monetary and fiscal policies in the effort to get the unemployment rate below 5 percent. The expansion of aggregate demand resulting from this policy will involve some upward pressure on the inflation rate and, assuming that the growth of the money wage rate does not increase to offset this, there will be some reduction in the real wage rate. This gives employers an incentive to expand output, and to do this they will expand employment. Suppose that over time the employment rate is reduced to 3 percent as a result of a rise in the inflation rate to 6 percent as shown by the point B. We then have the conventional Phillips curve conclusion that a lower rate of unemployment may be obtained at the cost of a higher rate of inflation. However, the critics assert that the decline in the unemployment rate obtained by this tradeoff will only be temporary. Labor will dis-

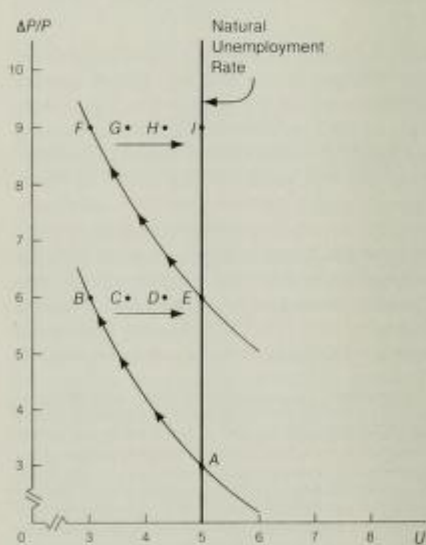


FIGURE 23-10

[20]Foremost among these are Phelps and Friedman. See E. S. Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium" in *Journal of Political Economy*, July–August 1968, Part II, pp. 678–711 and *Inflation Policy and Unemployment Theory*, W. W. Norton, 1972, Ch. 2; M. Friedman, "The Role of Monetary Policy," in *American Economic Review*, March 1968, pp. 1–17. For a relatively nontechnical comparison of the views of these economists with the trade-off view, see R. W. Spencer, "The Relation Between Prices and Employment: Two Views," in *Review*, Federal Reserve Bank of St. Louis, March 1969, pp. 15–21.

cover before long, especially if it has recently gone through another period of inflation, that the real wage has been reduced by inflation and will demand and receive the larger percentage increase in the money wage rate needed to offset the higher inflation rate. As this occurs, the real wage rate, which had been depressed by inflation, starts to return to its earlier level. The Phillips curve will shift upward to the right as the adjustment takes place, and as it does, assuming no further rise in the inflation rate, the unemployment rate rises from 3 to 5 percent as traced by the points *B*, *C*, *D*, and *E*. At this stage the unemployment rate has returned to its original level of 5 percent. The net result is a higher inflation rate and a higher rate of increase in the money wage rate but no change in the unemployment situation. The higher rate of inflation secures a temporary decrease in the unemployment rate, not the permanent decrease that is suggested by the tradeoff argument.

A further expansion of aggregate demand by forcing the inflation rate still higher will again provide a temporary reduction in the unemployment rate to the degree that wage increases again lag behind price increases. This could produce a movement from point *E* to point *F*, which parallels the earlier movement from point *A* to point *B*. However, this decrease in the unemployment rate will again be eliminated, according to the anti-tradeoff argument, as wage rates catch up with prices and wipe out the reduction in the real wage rate. The unemployment rate returns to 5 percent, as traced by the points *F*, *G*, *H*, and *I*. Again, the net result is a still higher inflation rate and an equally higher rate of growth of the money wage rate, but the unemployment rate is as it was.

If we think of a situation in which prices are rising not in intervals as above described but at a steady rate over a prolonged period of time, it is likely that wage increases will stop lagging behind price increases as labor comes to fully anticipate further price increases and to build into its labor contracts deferred wage adjustments to cover expected price rises. To the degree that discrepancies between expected and actual price and wage changes are eliminated, the reduction in the real wage via a rise in prices unmatched by a rise in wage rates no longer occurs, so that the reduction in the real wage rate, which is the means by which the unemployment rate is reduced, no longer occurs. The trade-off between higher prices and the unemployment rate disappears in the long run. In other words, there is no specific rate of price change related to a specific rate of unemployment when price changes are fully anticipated and reflected in wage rate changes. The unemployment rate moves to its equilibrium level and at this level is consistent with any rate of change of prices.

What this means graphically is that the Phillips curve becomes a vertical line like that erected in Figure 23-10 at the 5-percent rate of unemployment, that rate here being assumed to be the equilibrium rate of unemployment. Although expansionary policy through its effect on prices may be able to temporarily push the unemployment rate below this rate, the eventual catchup in wage rates will force unemployment back to this level. According to this view, there is no means by which we may choose to accept a higher rate of inflation in exchange for a lower rate of unemployment other than on a temporary basis. We will eventually end up with a rise in money wage rates equal to the rate of inflation and a return to the equilibrium unemployment rate shown here by the vertical line.

What is this equilibrium rate of unemployment toward which the system gravitates? It is the rate that corresponds with that real wage rate which equates the demand for and supply of labor, or it is essentially that rate of unemployment which corresponds with full employment as that term was defined in Chapter 14 on the classical theory.[21] If unemployment were less than this, there would be an excess demand for labor that would produce upward pressure on the real wage rate; if it were greater,

[21]See p. 232.

there would be an excess supply of labor that would produce downward pressure on the real wage rate. What the actual rate of unemployment will be at that real wage rate which equates the supply of and demand for labor, what Friedman calls the natural rate of unemployment, depends, in his words, on "the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on."[22] Thus, among other things, developments that improve the working of labor markets, such as greater mobility, more generally accessible information on job vacancies, and elimination of product and labor market monopolies, will tend to shift the curve to the left. However, according to these economists, it retains its vertical long-run shape for the reasons given earlier.

As we have seen, the record of the seventies to date has raised the question in the minds of some as to whether the Phillips curve relationship has broken down altogether, but there are those who maintain the position that not only does it exist but that it is of more than a temporary nature. Their view seems to be that, after any given unemployment rate has persisted long enough for the price-wage interactions to stabilize, the wage-rate increase will not be independent of that unemployment rate but will vary inversely with it in the way shown by the downward-sloping Phillips curve. This in turn leads to the conclusion that there is a durable trade-off that is denied by the other view. However, those who hold to this trade-off position grant that it is not nearly as stable or durable as they thought up to the late sixties. In place of the earlier belief that the downward-sloping Phillips curve's position remained unchanged over a period of many years, the common belief today of those who still believe there is a trade-off is that the curve is something less than vertical and that it does shift over time more than was earlier thought.

A CONCLUDING NOTE

The U.S. economy has experienced four sharp upward movements in prices since World War II: 1946–48, 1950–51, 1956–58, and the long period starting in 1965. Although the pronounced slowdown in the inflation rate during 1976, if sustained, will permit one to put a terminal date on this last period before many more years have passed, whether that will happen is highly uncertain.

What were the causes of these particular inflations? In this chapter we have looked into the principal theories advanced by economists to explain inflations in advanced economies. It is yet another matter to identify which theory or theories are applicable to any specific inflationary episode. The 1946–48 inflation following World War II and the 1950–51 inflation at the beginning of the Korean War are usually regarded as primarily demand-pull inflations because the economy was operating at high levels of employment and output in those years. The 1955–57 inflation was of a different nature. The economy then was operating with a good deal of excess capacity and demand pull is not the source of inflation under these conditions. Supply-side explanations were brought into the picture with some arguing that spontaneous increases in money wage rates were the major cause of the surge in prices. When one turns to the long inflation that started in 1965, he finds that excess demand appears to have been the major source in the early years. As of 1969

[22]M. Friedman, "The Role of Monetary Policy," op. cit., p. 8.

the unemployment rate was, relative to what followed, a remarkably low 3.5 percent. However, in the following years cost-push forces appear to have become important, but identification becomes a more difficult matter. In 1973–74 special factors like bad crops, dollar devaluation, and the OPEC cartel, come into play. In many cases the identification of causation presents perplexing, even impossible problems, but the closer economists can come to an understanding of the causal forces at work in any actual inflationary process, the more appropriate and effective can be the policy measures taken to control that inflation. An inflation may burn itself out on its own, if the inflationary process itself leads to a decrease in aggregate real demand. But to check an inflation before it has done the great damage that is characteristic of an inflation left on its own requires goverment intervention, and the types of intervention appropriate to demand-pull are not equally appropriate to cost-push. Restrictive monetary and fiscal policies are the standard remedies where the cause is clearly excess demand, but the same kind of policy cannot be so freely used where the cause is clearly cost-push. To curb cost-push inflation, measures such as stronger anti-trust enforcement and anti-union legislation may be considered, but such measures are not relevant to demand-pull. Since the policy measures to be employed in an attack on an existing inflation depend on the nature of the particular inflationary forces at work, the need for identification of these forces is as important as the need for an understanding of the inflationary process in general.

chapter 24

Incomes Policy and Fiscal Policy

Since 1946 it has been the responsibility of the federal government to work toward the achievement and maintenance of maximum employment, price-level stability, and a high rate of economic growth. The Employment Act of that year called on the government for the first time to

> use all practicable means consistent with its needs and obligations and other essential considerations of national policy . . . to coordinate and utilize all its plans, functions, and resources for the purpose of creating and maintaining, in a manner calculated to foster and promote free competitive enterprise and the general welfare, conditions under which there will be afforded useful employment opportunities, including self-employment, for those able, willing, and seeking to work, and to promote maximum employment, production, and purchasing power.

The wording of the act, it will be noted, explicitly covers only the single goal of maximum employment. However, as interpreted by numerous executive department statements and actions in which the Congress has concurred, the goals are generally understood to include price stability and satisfactory economic growth as well as the one directly stated.

Although the Employment Act is a landmark in economic legislation, its passage in 1946 does not mean that the federal government was not previously aware of a responsibility in the area to which the act refers. Still, there is a great difference between a mere awareness of a responsibility to work toward the goal of maximum employment and legislation that specifically directs that "all practicable means" be used to achieve this goal. The 1946 act does in this sense replace what at best was a vague and undefined sense of responsibility with an obligation that is somewhat more precisely defined. In this regard, it is interesting to speculate on what differences there would have been in the actions taken by the federal government during the years of the Great Depression if the Employment Act of 1946 had appeared instead in 1926. Perhaps the Great Depression would have been much less great!

It may well be argued, however, that an act of this kind could not have come in 1926 and that it did come in 1946 only because, over the decade preceding 1946, the conviction had grown that the federal government not only had the responsibility to work toward the attainment

of the specified goals but could make a truly significant contribution to the attainment of these goals. The power of the federal budget to eliminate unemployment had been revealed in dramatic fashion during the years of World War II. The conviction that the budget could and should be used as a means of helping to provide employment opportunities in peacetime had come to be generally held by 1946. This conviction stands in sharp contrast to one held by many before World War II, a conviction that underlay what one writer has described as the *moralistic economics* of those earlier years.[1] Those who accepted this moralistic economics adopted a "boom and bust" attitude toward business fluctuations. In a word, the causes of the bust were found in the excesses and maladjustments of the boom—in the speculation, unsound uses of credit, high living, and the like. The severity of the downturn also depended on the severity of the "economic sins" committed during the boom. Most importantly, because the moralistic economics did not recognize that government could, with the necessary tools in hand, contribute substantially to the stabilization of the economy, it could offer little help toward eliminating "boom or bust" except to preach an economic piety that might somehow reduce that instability. This kind of thinking appeared to reach a high point in the late twenties and early thirties, and to combat it was undoubtedly one of Keynes's purposes in writing the *General Theory*.

As we noted in the first chapter, Keynes's book was a success with few equals in the history of economics. Without it and its impact on the world of practical affairs, we would probably not have had the Employment Act of 1946. And had there been no Employment Act of 1946, it is doubtful that we would have had many of the federal government actions taken to attain high employment and promote economic growth during the years since World War II. For with the act came the establishment of the Council of Economic Advisers in the executive branch and the Joint Economic Committee in the legislative branch. And through them and their research studies, committee hearings, and the annual Economic Report of the President came a better understanding on the part of the Congress, business, and the general public as to what actions by the federal government might help in achieving these goals and what actions might work in the opposite direction.

Of course, throughout the thirty-plus years since the passage of the act there have been differences of opinion as to the degree of responsibility that government should assume in trying to achieve these goals. As might be expected, these differences even go back to the debate that preceded the passage of the act in 1946. The original bill known as the Murray Bill, after Senator Murray of Montana, was formally submitted in 1945 as the "Full Employment Bill." Most important among its various provisions, this bill declared in effect that the United States, as a nation and a government, had the responsibility to step in and provide jobs for all those seeking work and able to work who were unable to find work. In the debate on this bill, the position of those who favored provisions that imposed less responsibility on government as a guarantor of jobs was the one that prevailed. The final bill signed into law by President Truman nowhere contained the term "full employment," although the original Murray Bill had repeatedly used that term but oddly had in no place defined it. As noted in the above excerpt from the bill that was passed, the government's responsibility was limited to the less specific objective of promoting "maximum employment."

What was true at the time of the debate on the Murray bill is true today; there are those who insist that the government's responsibility goes far beyond that imposed on it by the 1946 act. Prominent among these are Senator Humphrey who with Representative Hawkins of California has introduced "The Full Employment and

[1]See J. M. Culbertson, *Macroeconomic Theory and Stabilization Policy*, McGraw-Hill, 1968, pp. 371–73.

Balanced Growth Act." This bill not only imposes on the government the responsiblity for "full employment," but sets the time period within which full employment as defined in the bill is to be achieved. The 1976 draft of the proposal required that the adult unemployment rate be reduced to 3 percent within four years after passage of the act. The detailed actions that presumably would achieve this goal are set forth in the bill, including the assignment to the federal government of the role of "employer of last resort." The federal government would hire, at prevailing wage rates, whatever members of the labor force were unable to find employment elsewhere. As did the Murray bill thirty years earlier, the Humphrey-Hawkins bill generated strong opposition, not only from the business sector but also from some respected economists. Among other objections, they assert that such an ambitious goal cannot be achieved without a disastrous inflation, a development which in turn may lead to a recession as bad as or worse than that of 1973–75. The supporters see problems to be resolved, but, needless to say, they see no disaster of this or any other kind if all aspects of the detailed program are implemented.[2] Whether some draft of this bill will eventually become law only time will tell. The 1976 draft is a watered-down version of an earlier one that a majority of economists believed was unrealistic. Perhaps again a compromise will be reached in the form of an act that imposes on the federal government substantially greater responsibility than does the Employment Act of 1946, but still less responsibility than would the Humphrey-Hawkins bill in its 1976 version.

However modest or ambitious the economic goals that acts of Congress direct the government to work toward, specific actions must be taken in the effort to attain them. Each such action represents the implementation, in whole or in part, of a particular policy—policy itself best being thought of as the specification of a prescribed course of action that is intended to achieve or contribute to the achievement of certain goals. Just as economists refer in the broadest of categories to the three basic macroeconomic goals of full employment, price stability, and satisfactory economic growth (or four such goals, if we include balance-of-payments equilibrium), they similarly refer in equally broad categories to monetary policy, fiscal policy, and incomes policy as the basic types of policy that are employed in working toward the achievement of one or more of the specified goals. However, our macroeconomic goals, quite apart from the problem of what is really meant by such loose terms as "full employment," "price stability," and "satisfactory growth," cannot really be held within the confines of the three or four terms that make up the standard list. There are at least two more major goals that should be identified: the goal of economic justice, the principal characteristic of which is an "equitable" distribution of income, and the goal of economic freedom, characterized by the right of every man to change jobs, join a labor union, enter a business, own property, purchase the goods he wants, and do endless other things.

Similarly, our macroeconomic policies actually include some that are quite beyond what is covered by monetary, fiscal, and incomes policy in the widest sense of those terms. Although these three are the policies we understandably hear the most about, we should note that policies such as labor, agricultural, antitrust, tariff, public housing, conservation, and others even less familiar are all relevant in some way to the multiple macroeconomic goals we pursue in the real world.

[2]For some 800 pages of views of various economists, labor, business, and other groups, see *Full Employment and Balanced Growth Act, 1976*, Hearings before the Subcommittee on Employment, Poverty, and Migratory Labor, Senate Committee on Labor and Public Welfare, May 1976. See also *An Economic Analysis of the Full Employment and Balanced Growth Act of 1976*, Congressional Budget Office, Congress of the United States, May 1976, reprinted as pp. 800–837 of the above hearings, and M. C. Barth, *The Full Employment and Balanced Growth Act of 1976: An Analysis and Evaluation*, Institute for Research on Poverty, Discussion Papers, Madison, Wisconsin, 1976.

A treatment of anything like a full list of macroeconomic goals and a full list of policies relevant to the pursuit of these goals would require an entire book or even several books. In our restricted coverage, we will bypass such goals as economic justice and economic freedom, not because they are unimportant but because of limited space and because we have not developed in even the crudest form a theoretical framework needed for an analysis of policies appropriate to these goals. From earlier chapters we do have a theoretical framework adequate for an analysis of certain policies aimed at the specific goals of full employment, price stability, and economic growth, and it is to these three goals that we will limit our attention. We will also limit our attention to three types of policies: incomes, fiscal, and monetary, with emphasis on the latter two.

Monetary and fiscal policy have already been introduced in various contexts in earlier chapters. Thus, in developing the simple Keynesian model of income determination in Part 2, we occasionally mentioned specific monetary and fiscal policies designed to raise the levels of real income and employment. Similarly, in the preceding chapters of Part 5, we referred at several points to monetary and fiscal policies relevant to the problems of the business cycle, economic growth, and inflation. However, all such references were incidental to the development of the basic theory of income and employment and of certain theories of the cycle, of growth, and of *IS-LM* framework that in a depression fiscal policy will be more effective than monetary policy in raising the income level. But to formulate specific fiscal policies to meet these conditions we need much more information than a theoretical framework like *IS-LM* can give us. Policies formulated for the actual economy must face the many practical problems that arise out of complex political, social, and economic institutions, procedures, and practices. While all these must be considered, it is also to be emphasized that the theoretical framework remains at the foundation of rational policy making and is absolutely indispensable to it. For without an underlying theory a policy-maker would have no notion of what consequences to expect from any proposed policy and thus no way of knowing whether the policy in question would be a help or a hindrance in achieving the goal that is the very reason for adopting the policy. In a word, there can be no intelligent policy in the absence of a well-reasoned theoretical framework on which to base policy. As we proceed in these two chapters, it will become apparent both how the theoretical framework developed in earlier chapters lies at the bottom of policy formulation and why much more than just that theoretical framework is required for policy formulation.

Although nobody among those who believe that there is a role for incomes, fiscal, and monetary policy questions that the various policies must be coordinated if we are to get the best results, it is very convenient and a

out the economy. How changes in specific prices and wage rates are worked out in response to changes in aggregate demand is a matter of market forces; monetary and fiscal policy only produce the changes in demand that set into motion a process involving these other changes. However, the changes in prices and wage rates brought about by the operation of these market forces may be such at times as to severely qualify the usefulness of demand-oriented monetary and fiscal policies. For example, an expansion of aggregate demand needed to reduce unemployment may be engineered through monetary and fiscal policy. But if reducing unemployment to an acceptably low rate involves an unacceptably high rate of inflation, then the use of monetary and fiscal policies becomes an unacceptable means of meeting the unemployment problem. This trade-off between the unemployment rate and the inflation rate was discussed at length in Chapter 23, but we did not go into the question of whether there is another policy available that can meet this conflict among goals. In other words, we did not ask whether we can use monetary and fiscal policy to expand aggregate demand sufficiently to move an economy suffering unemployment toward its full-employment position and then use another policy to limit the wage and price inflation that tends to result as we approach the full-employment position. There are indeed various courses of action that can be taken in order to hold wage and price advances below those that would result from the working of unimpeded market forces, and it is these various courses of action that constitute what is usually meant by the term *incomes policy*.

Of these courses of action, probably the weakest is "jawboning" or "moral suasion" in which government seeks to talk unions and business into restraining their wage demands and price increases. The strongest are wage and price controls in which legal increases in wages and prices occur only on the approval of government agencies that administer the wage and price controls and not merely because market forces are working in that direction. Between these extremes are other courses of action such as the use of wage and price guideposts, an approach followed in the United States in the mid-sixties and then again from 1971 to 1974.

The Carter administration in its first few months in office came out flatly against resort to outright wage and price controls and also expressed reluctance to make use of wage and price guideposts. What was established was a voluntary, cooperative program in which representatives from management, labor, and the federal government would form a monitoring unit. The president of the AFL-CIO and the chairman of General Electric Corporation were to head this panel. The Secretary of the Treasury would represent the government. An existing agency, the Council on Wage and Price Stability, would be given additional power and would step up its studies of wage and price increases. It would hold hearings and issue warnings of impending shortages. The Council could also request confidential information from companies and unions in carrying out its responsibilities. As a purely voluntary program, the prices charged by companies and the wage increases obtained by unions would be subject to no legal restraint. However, as with such programs in the past, the expectation is that the pressure of public opinion will be sufficient to prevent companies and unions from pushing through increases in prices and wages which have been judged to be excessive. Of course, as with earlier voluntary programs, there is always the threat that the voluntary program will be replaced with a mandatory program if the voluntary program is unsuccessful. This would mean a return to the kind of incomes policy employed in this country several times in the past.

No one can say whether the United States will again resort to that kind of an incomes policy, but it is surely more than a remote possibility. And if a step is taken on incomes policy beyond the purely voluntary program established in the early months of the Carter administration, it will most likely be toward some system of wage and price guideposts. In the following pages we will outline the economic rationale for this kind of

out the economy. How changes in specific prices and wage rates are worked out in response to changes in aggregate demand is a matter of market forces; monetary and fiscal policy only produce the changes in demand that set into motion a process involving these other changes. However, the changes in prices and wage rates brought about by the operation of these market forces may be such at times as to severely qualify the usefulness of demand-oriented monetary and fiscal policies. For example, an expansion of aggregate demand needed to reduce unemployment may be engineered through monetary and fiscal policy. But if reducing unemployment to an acceptably low rate involves an unacceptably high rate of inflation, then the use of monetary and fiscal policies becomes an unacceptable means of meeting the unemployment problem. This trade-off between the unemployment rate and the inflation rate was discussed at length in Chapter 23, but we did not go into the question of whether there is another policy available that can meet this conflict among goals. In other words, we did not ask whether we can use monetary and fiscal policy to expand aggregate demand sufficiently to move an economy suffering unemployment toward its full-employment position and then use another policy to limit the wage and price inflation that tends to result as we approach the full-employment position. There are indeed various courses of action that can be taken in order to hold wage and price advances below those that would result from the working of unimpeded market forces, and it is these various courses of action that constitute what is usually meant by the term *incomes policy*.

Of these courses of action, probably the weakest is "jawboning" or "moral suasion" in which government seeks to talk unions and business into restraining their wage demands and price increases. The strongest are wage and price controls in which legal increases in wages and prices occur only on the approval of government agencies that administer the wage and price controls and not merely because market forces are working in that direction. Between these extremes are other courses of action such as the use of wage and price guideposts, an approach followed in the United States in the mid-sixties and then again from 1971 to 1974.

The Carter administration in its first few months in office came out flatly against resort to outright wage and price controls and also expressed reluctance to make use of wage and price guideposts. What was established was a voluntary, cooperative program in which representatives from management, labor, and the federal government would form a monitoring unit. The president of the AFL-CIO and the chairman of General Electric Corporation were to head this panel. The Secretary of the Treasury would represent the government. An existing agency, the Council on Wage and Price Stability, would be given additional power and would step up its studies of wage and price increases. It would hold hearings and issue warnings of impending shortages. The Council could also request confidential information from companies and unions in carrying out its responsibilities. As a purely voluntary program, the prices charged by companies and the wage increases obtained by unions would be subject to no legal restraint. However, as with such programs in the past, the expectation is that the pressure of public opinion will be sufficient to prevent companies and unions from pushing through increases in prices and wages which have been judged to be excessive. Of course, as with earlier voluntary programs, there is always the threat that the voluntary program will be replaced with a mandatory program if the voluntary program is unsuccessful. This would mean a return to the kind of incomes policy employed in this country several times in the past.

No one can say whether the United States will again resort to that kind of an incomes policy, but it is surely more than a remote possibility. And if a step is taken on incomes policy beyond the purely voluntary program established in the early months of the Carter administration, it will most likely be toward some system of wage and price guideposts. In the following pages we will outline the economic rationale for this kind of

incomes policy and review the U.S. experience with it in the mid-sixties and early seventies.

The Rationale for Wage-Price Guideposts

In an economy in which the money wage rate rises faster than the productivity of labor, the tendency is for the price level to adjust upward by the amount necessary to cover the difference. This follows from the basic profit maximization rule introduced earlier: $W = P \cdot \text{MPP}$. Given the money wage rate, employers hire that amount of labor for which the price of output times the marginal physical product of labor is equal to the money wage rate. Thus, if W rises by 5 percent in any year but the MPP of labor also rises by 5 percent in that year, W/MPP (i.e., the marginal cost) remains unchanged, and an unchanged P is still consistent with profit maximization. On the other hand, a percentage rise in W that exceeds the percentage rise in the MPP of labor indicates that a higher price level is necessary for profit maximization. In this case, W/MPP has increased, and, unless prices rise to allow for the rise in cost, the rise in the wage rate occurs at the expense of the profit margin. The tendency in practice is for the price level to adjust by the amount necessary to cover any difference between the percentage changes in W and in the MPP of labor.

The data for the postwar years on labor productivity, compensation per employee work-hour, and price level attest to this tendency. Over these years, prices have moved roughly in line with the difference between the change in labor productivity and the change in compensation per employee man-hour. The productivity increase in 1951 was 2.9 percent, the increase in compensation per employee man-hour was 9.8 percent, and the price level increase was 7.3 percent. Ten years later in 1961, the figures were 3.3 percent for productivity increase, 3.5 percent for compensation increase, and 0.6 percent for the price-level increase. For more recent years, consider 1972 and 1976. In 1972, the increase in productivity was 2.9 percent, in compensation 5.7 percent, and in the price level 3.6 percent. In 1976, the corresponding figures were 4.0, 7.7, and 5.2 percent.[3]

While there is more behind price-level movements than the relative changes in labor productivity and labor compensation, there is little doubt that the goal of reasonable stability in the price level cannot be attained in an economy in which the compensation of labor per man-hour is pushed up at a rate well in excess of the rate of increase in productivity per man-hour. This is why a money wage rate that rises at the same rate as "labor productivity" is popularly described as "noninflationary," or consistent with price-level stability, and one that rises at a more rapid rate as "inflationary," or inconsistent with price-level stability.

Because this relationship is indicated by the theory and supported by the postwar data, the Council of Economic Advisers, in its January 1962 *Annual Report*, officially advanced for the first time the idea of general wage and price guideposts for price stability.[4] In its briefest form, the general wage guidepost for annual increases in total compensation per employee man-hour is that such increases in percentage terms should not exceed the trend of labor productivity.[5] If so limited, increases in labor com-

[3]Figures for 1976 are from *Monthly Labor Review*, July 1977, and *Survey of Current Business*, April 1977; all other figures are from *Annual Report of the Council of Economic Advisers*, January 1977, p. 229.

[4]*Annual Report of the Council of Economic Advisers*, Jan. 1962, pp. 185–90. The genesis and principles of the guideposts are reviewed in the *Annual Report*, Jan. 1967, pp. 120–34. For subsequent statements, see the *Annual Report*, Feb. 1968, pp. 120–28, and Jan. 1969, pp. 118–21. The development of the guideposts as seen in the *Annual Reports* from 1962 to 1967 is examined in R. E. Slesinger, *National Economic Policy—The Presidential Reports*, Van Nostrand, 1968, pp. 100–27.

[5]Our model compared the "wage rate" with the "marginal physical product" of labor. Since, in practice, the cost of labor per hour includes all kinds of fringe benefits as well as the hourly wage rate, our use of the term "wage rate" must here be broadly interpreted as total compensation per employee man-hour. Secondly, since the productivity change in any one year can be influenced by short-run transitory factors, the Council used the trend productivity, which is the annual average percentage change in output per man-hour during the latest five years.

pensation will be consistent with stable prices and an unchanged distribution of income between labor and others.

An arithmetic example will show how these results follow. Assume that a worker is paid $3.00 per hour, or $120.00 for a forty-hour week, and that output per worker per week is 200 units. Output per man-hour is 200 divided by 40, or 5 units. Labor cost per unit of output is $120 divided by 200, or $0.60, or, what is the same thing on an hourly basis, $3.00 divided by 5 equals $0.60. If as a result of a technological advance, better capital equipment, or any other reason, output rises by 3 percent to 206 units per week for the same 40 hours of labor, output per man-hour is also higher by 3 percent, or is now 5.15 units. If the worker receives an increase in compensation equal to the increase in productivity, he will get $3.09 per hour, or $123.60 for the forty-hour week. Labor cost per unit of output, however, remains unchanged at $0.60, or $3.09 divided by 5.15. If the price at which the product is sold remains unchanged, the difference between price per unit and labor cost per unit, which is the amount available for payment to others including stockholders, will also remain unchanged. However, with a 3-percent increase in the number of units produced and sold, the total amount remaining after labor cost is also higher by 3 percent. The wage guidepost thus indicates that a rise in the wage rate equal to the gain in productivity is consistent with a stable price for the product and a percentage increase in the earnings of the nonlabor factors equal to that received by labor.

Because productivity gains vary widely by industry and firms, the guidepost approach requires a price as well as a wage guidepost. The general price guidepost, in its simplest form, is that those firms that grant wage increases equal to the national trend productivity but experience a rise in productivity greater than the national trend productivity should reduce prices by an amount to reflect this difference. Conversely, those firms that experience an increase in productivity smaller than the national trend productivity, but nonetheless grant wage increases equal to the national trend productivity, would appropriately raise prices to cover that portion of the compensation increase that is not matched by the productivity increase. While some prices would thus fall and others rise, the overall result should be approximate price stability.

If a guidepost program is to have any chance of success in providing price stability, labor in general must settle each year for an increase in compensation no greater than the wage-guidepost figure announced for that year, and firms in general must set prices that correspond to the price guidepost. The failure of firms to observe the price guidepost can lead to a rising price level in the same fashion as the failure of labor to limit its demands to the wage guidepost. The exercise of market power to raise prices to exploit what firms regard as favorable demand conditions for their output differs in kind but not in result from the exercise of market power by organized labor to raise wage rates to exploit favorable demand conditions in the market for labor.

U.S. Experience with Wage-Price Guideposts

As noted above, the original official proposal for guideposts in the United States came in the January 1962 *Annual Report of the Council of Economic Advisers*. However, at that time no specific trend productivity was set forth. Two recessions, 1957–58 and 1960–61, had occurred in the preceding five years, and it was difficult to identify the trend productivity for this period. Although no specific figure was stated in the next two *Reports*, in the 1964 *Report* the subsequently well-known figure of 3.2 percent appeared as the latest figure in a column labeled "trend productivity." In that year the 3.2-percent figure came to be the recognized general guidepost for wages. In following *Reports*, the Council gave increasingly clear

indications of what it regarded as the trend of productivity. Thus in the January 1966 *Report*, the Council specifically recommended that the general wage guidepost be 3.2 percent for that year.

In the very same year there began a process that in effect temporarily marked the end of guideposts. Until mid-1966 the guideposts had been observed reasonably well by labor and business. On the part of labor, this to some degree may be attributed to a level of unemployment that averaged more than 5 percent during 1962–65. With increases in labor compensation in most cases limited to the guidepost figure and with business pricing decisions generally in line with the price guidepost, the consumer price index rose less than 2 percent per year during 1962–65. Then the rapid upward surge in the economy in 1966 produced a fall in unemployment to 3.8 percent, and the resulting tightness in the labor market subjected the 3.2-percent wage guidepost to pressures unknown in the preceding years. More and more labor settlements were reached at figures in excess of the guidepost. Although it is more difficult to generalize on the side of prices, price decisions of firms with price discretion also appeared, more frequently than in preceding years, to be inconsistent with the decisions called for by the price guidepost. The consequence was that consumer prices rose by 3.3 percent during 1966, which meant that in January 1967 the Council was effectively barred from announcing a wage guidepost for 1967 equal to the trend productivity. A figure in the neighborhood of 3.2 percent would, on the average, have provided the worker with an increase in compensation no more than sufficient to cover the rise in living costs and therefore would have permitted no increase to reflect the rise in his productivity. This would be clearly unacceptable to the unions. For a case in point, a story in *Business Week* of November 26, 1966, was headlined "Unions Call Five Percent a Minimum for 1967." On the other hand, for the Council to have officially announced a guidepost figure that was adjusted upward by the amount needed to allow for all or part of the rise in living costs would have been inconsistent with the basic guidepost objective of preventing a rising price level. For example, a guidepost of 5.5 percent made up of 3 percent for productivity gain and 2.5 percent for the rise in the cost of living would have amounted to the adoption of a guidepost that officially accepted approximately a 2.5-percent rate of inflation. Because the guideposts had been popularized over the years just preceding as a means of maintaining price stability, a wage guidepost that actually sanctioned inflation, however moderate, would not have been favorably received by the general public. The conditions being what they were, the Council reacted to the further 3-percent increase in consumer prices in 1967 as it had to the similar rise the year before: as in its 1967 *Report*, no guidepost figure was announced in its 1968 *Report*. For the time being, guideposts were dead.

That they would remain dead was assured by the new administration that took power in January 1969. President Nixon's Council of Economic Advisers and others in high positions wasted little time in making known the President's and their opposition to so-called "incomes policies" in all forms from "jawboning" to guideposts to price and wage controls. As reviewed in the preceding chapter, for more than two years the administration steadfastly limited its attack on inflation to what could be done through monetary and fiscal policies. Then with these policies apparently showing little success, in a sudden surprising about-face, on August 15, 1971, the President adopted the most extreme form of the kind of policy that had been denounced uninterruptedly for about two years. On this date he imposed a 90-day freeze on practically all prices and wages, the ultimate form of intervention in the marketplace. The 90-day freeze was followed by the establishment of a system of wage and price guideposts with mandatory rules or standards and with compliance required by law. This sys-

tem was in effect from November 13, 1971, to January 11, 1973. On the latter date, the system of guideposts was modified to a predominantly quasi-voluntary form under which the standards were to be self-administered and voluntary behavior consistent with these standards was to be expected. In the President's words, the system would be "as voluntary as possible . . . but as mandatory as necessary." These three periods and the control system in effect during each were known as Phase I, Phase II, and Phase III.

With the abrupt rise in the cost of living during the early months of 1973, primarily the result of the extreme increase in food prices, Phase III came under heavy attack by the public and by Congress. The administration had originally turned to Phase III as a step toward a return to the pre-August 1971 system of no wage and price controls, but the sharp rise in prices that occurred in 1973 suggests, to say the least, that this transition had begun too soon. In the face of price increases that grew greater month by month, the President again reversed direction. On June 13, 1973, he imposed the second price freeze in less than two years. This ran for two months, and was succeeded on August 12 by Phase IV, a system that was stricter in certain ways than Phase II for the industries it covered. However, this system at the same time incorporated a new strategy to achieve a return to free markets: decontrol of selected industries where that was deemed appropriate. Decontrol was authorized for some industries on reaching an agreement that the firms in those industries would expand output or capacity or limit exports and in this way ease price pressures in the domestic market by increasing supplies. Some other industries were decontrolled on the grounds that there had been no build-up of cost pressures in those industries. In such cases decontrol would not lead to a spurt in prices and therefore no advantage would be gained by maintaining controls. Thus, while there was a move toward a tightening of controls in some areas, there was also a move toward decontrol in others. In any event, as things developed, the system known as Phase IV came to an end on April 30, 1974, the expiration date of the Economic Stabilization Act whose extension the administration did not request (except in the health and construction fields) and the Congress did not choose to provide on its own. For all practical purposes, after April 30, 1974, prices and wages in the U.S. economy were again permitted to reach their free-market levels as had been the case before August 15, 1971, when the 32-month episode of controls began.

From its beginning to its end, the objective of the control program was, of course, simply to hold down the rate of inflation, recognizing that this could only be accomplished if both wage-push and profit-push forces could be restrained by the program. During the first freeze or Phase I in 1971 the annual rate of inflation as measured by the consumer price index was slowed by more than one-half from the prefreeze rate, or from about 4.0 to 1.9 percent. The rate did not fall to zero because prices of some consumer items were unaffected due to exemptions and inapplicability of the controls. With the end of the freeze, it was conceded by the authorities that the rate of price advance would rise above that during the freeze period, but it was hoped that, apart from a temporary surge, it would be held below the prefreeze rate. In any event, this meant that the new wage guidepost designed to go into effect with the end of the freeze could not be set at 3 percent or approximately equal to the productivity increase. Labor would justifiably be up in arms. The problem here was the same as that during 1967 and 1968: either no guidepost or one that sanctioned some amount of inflation. In 1971 the latter option was chosen—a wage guidepost was adopted that allowed for some inflation as the price that had to be paid to get into operation a system of guideposts that might over time be able to reduce the rate of inflation. The Pay Board, part of the administrative machinery set up for Phase II, was given the responsibility of laying down general policy

governing maximum pay increases or, in other words, setting the wage guidepost. As Phase II went into effect in November 1971, the Board set a rate of 5.5 percent as the basic standard for new labor contracts and adjustments. With this initial percentage almost twice as large as the productivity advance, it was one that plainly incorporated an allowance for some amount of inflation.

Side by side with the Pay Board, a Price Commission was established to administer prices. It set forth the general rule that no price might be increased beyond the ceiling price established for the freeze period (which in turn was the highest price at or above which a "substantial volume" of transactions had taken place in the 30 days preceding the freeze). However, price increases would be permitted to reflect increases in costs subject to a profit margin limitation to be noted below. The goal and the expectation was that the rate of average price increase would be held to 2.5 percent, a goal that, if realized, might bring the rate of inflation down to 2 to 3 percent by the end of 1972. To hold the rate of price increases down to 2.5 percent at a time when the prefreeze rate had been far above this appeared quite ambitious. However, this percentage is quite consistent with that set as the wage standard by the Pay Board. If the average rate of increase in labor compensation could be held to the 5.5-percent standard, labor cost per unit of output would rise at about a 2.5-percent rate on the basis of a labor productivity trend of about 3 percent per year. If nonlabor costs—notably depreciation, indirect business taxes, and interest, which together make up about one-fourth of total costs—also rose per unit of output at a rate of about 2.5 percent, permitting firms a corresponding 2.5-percent rate of price increase would protect their existing profit margins and maintain an unchanged division of income between labor and profit shares.

What does the record show? From November 1971 to February 1972 there was the bulge in prices that was expected to follow the end of the freeze. The consumer price index rose at a 4.8-percent annual rate during these months. However, over the balance of 1972, it rose at only a 3-percent annual rate. Just as certain prices were promptly adjusted upward to correct for the inequities of the freeze, so too were some wage rates. From November 1971 to February 1972, average hourly earnings of workers in the private nonfarm economy rose at an annual rate of 9.5 percent, but during the balance of 1972 the rate was 5.6 percent. While the rates of both wage increase and price increase during Phase II as a whole exceeded the goals set, both rates still showed an impressive decline relative to those in the pre-control period.

Under Phase III the President established as a goal a rate of inflation of 2.5 percent or less by the end of 1973. Initially at least, the general standards for wages and prices established under Phase II were retained and, as noted earlier, these were consistent with the 2.5-percent rate of price advance set forth as the goal. However, as of early 1973, it was clear that the favorable record achieved under Phase II would not be repeated under Phase III. Farm prices had not been covered under the control program from the beginning because of special difficulties faced in this area. A rapid rise in food prices began in late 1972 and accelerated in early 1973. For the six months from October 1972 to April 1973 the food component of the consumer price index increased by 9.3 percent. Food is, of course, a major item in the consumer's budget, and food prices are the prices with which he is in close and continuous contact. Although other prices rose no more rapidly than they had earlier in 1972, the consumer's impression of overall price movements is most influenced by those that show the most dramatic changes. What is at issue is this: The 5.5-percent wage guidepost cannot be maintained unless the working people generally believe in its equity, and such a belief cannot survive in the face of food prices rising at the rate noted, even though that rise may be cor-

rected before many months have passed. After looking at the record of food prices over the preceding few months, George Meany noted in late February 1973 that a wage guidepost of 7.5 percent rather than 5.5 percent might be in order. This kind of attitude was reinforced by looking ahead to the higher prices of imports that would follow from the 10-percent devaluation of the dollar that same month.

The year 1973 was an especially heavy one for labor negotiations—contracts were worked out covering 4.7 million workers compared with 2.8 million workers in 1972. In some new contracts, increases of more than 5.5 percent could be justified on "catch-up" or other grounds, but if a general pattern of increases much in excess of this rate were to occur, pressures greater than anticipated would be exerted on the price level. For the private nonfarm economy, the increase in average hourly earnings in 1973 turned out to be 7.8 percent in comparison to 5.6 percent from February 1972 to the end of that year. With the 7.8 percent increase in hourly earnings, the Phase III goal announced in January 1973 of reducing the rate of inflation to 2.5 percent by the end of the year could not be realized. At the end of 1973, prices were rising at about three times the rate that had been set as a goal. Then, due to special factors, especially the skyrocketing of oil prices, the price level increased at an annual rate greater than 15 percent during the first third of 1974, the last four months of the control program.

Under the standards established for Phases II, III, and IV, of the program, higher labor costs as well as higher costs of other kinds could, under most conditions, be passed along in higher prices, and these increases in prices could be well above 2.5 percent if actual increases in costs were sufficiently large. The basic device designed to prevent prices from rising more than necessary to cover firms' higher costs was the standard that firms' profit margins (profits before taxes as a percentage of sales) should not exceed those in a base period. This meant that higher costs did not automatically justify higher prices; they did so only if prices could be raised without exceeding base-period profit margins. In the more stringent final phase of the controls program, Phase IV, price increases in manufacturing and service industries were limited to the dollar and cents amount of allowable cost increases, with no add-on permitted to maintain the base period margins.

Looking back over the actual course of wages and prices during the almost three years of Phases I–IV, some economists maintain that the actual record of price and wage rate advance was little influenced by the fact that a control program was in operation. They grant as they must that there was an impressive slowing in the rate at which wage rates and prices rose during the first year under Phase II. However, they assert that this may simply have been the effect of the considerable slack in the economy that remained from the 1970 recession. Apart from this, it is usually the first year or so of a prolonged period of controls that shows the best results. There will often be a degree of support and cooperation by labor and business at first, but as the distortive and discriminatory effects of controls, some of which are unavoidable, begin to be felt, support is replaced by opposition. Also, as time passes, circumvention of the regulations becomes more common as the public works out ways in which this can be accomplished. As the anticontrol people see it, the inflation record over the full life of the program would have been little different if there had never been a guidepost program at all. And there are, of course, defenders of the program who not only give major credit to the Phase II wage and price guideposts for the favorable results of 1972, but also argue that the results in 1973 would have been much better if Phase II had not been replaced by the much weaker Phase III. In reply, the other side contends that no control system of this kind could effectively cope with the shortage situation that had developed in 1973 with many industries operating at capacity. There were then very rapid increases

in prices of raw commodities that were entirely outside the control system from its inception. And then with 1974 came the staggering increase in the price of imported oil, something entirely beyond the reach of the control system.

The question of the overall effects on price and wage inflation of the 32 months of experience with controls has been much debated.[6] Still it appears that no amount of debate can produce a final judgment. The major difficulty is that there is no way of accurately simulating what would have occurred in the areas of wage rates and prices over these years if there had been no control program at all. It does not automatically follow that a program established for the specific purpose of restraining advances of wage rates and prices will produce the net result of lesser increases than would otherwise occur; the opposite result is also possible. However, quite apart from the question of the efficacy of the program in practice, there is little question of the continuing validity of the principle underlying the wage guidepost that is the heart of any such program. It is safe to say that the following statement made by the Council of Economic advisers during the Johnson administration would be accepted without qualification by the Councils of the three administrations that have so far followed: "The only valid and noninflationary standard for wage advances is the productivity principle. If price stability is eventually to be restored and maintained in a high-employment U.S. economy, wage settlements must once again conform to that standard."[7] The question thus becomes whether conformance with that standard can be achieved without reliance on wage and price controls. Even conservative economists have their doubts on this. Arthur F. Burns who takes a quite conservative position on most economic questions summed up the problem in the following words in a statement made as chairman of the Board of Governors of the Federal Reserve System before a Senate committee in February 1973.

> The performance of the American economy in recent years, as well as that of other industrialized nations, has persuaded me that there is a need for legislation permitting some direct controls over wages and prices. I do not think that resort to such controls will be required all, or even much, of the time. However, the structure of our economy—in particular, the power of many corporations and trade unions to exact rewards that exceed what could be achieved under conditions of competition—does expose us to upward pressures on costs and prices that may be cumulative and self-reinforcing.[8]

[6]For some analyses of the operation and effectiveness of parts or all of the 1971–74 controls program, see the following: *Annual Report of the Council of Economic Advisers*, January 1974, Chapter 3, pp. 88–109; "Two Years of Wage-Price Controls," a series of papers in the *American Economic Review*, May 1974, pp. 82–104; B. Bosworth, "Phase II: The U.S. Experiment with an Incomes Policy," in *Brookings Papers on Economic Activity*, 2, 1972, pp. 343–83; R. J. Gordon, "The Response of Wages and Prices to the First Two Years of Controls," in *Brookings Papers on Economic Activity*, 3, 1973, pp. 765–78; D. J. Mitchell "Phase II Wage Controls," in *Industrial and Labor Relations Review*, April 1974, pp. 351–75; D. Q. Mills, "Some Lessons of Price Controls in 1971–73," in *Bell Journal of Economics*, Spring 1975, pp. 3–49, and "Recent Experience with Wage and Price Controls," in *Sloan Management Review*, Fall 1974, pp. 48–57; R. E. Azevedo, "Phase III—A Stabilization Program That Could Not Work," in *Quarterly Rewew of Economics and Business*, Spring 1976, pp. 7–21; D. Robinson, "Wage-Price Controls and Income Policies," in *Monthly Labor Review*, March, 1974, pp. 34–39; and A. R. Weber, "Making Wage Controls Work," in *The Public Interest*, Winter 1973, pp. 28–40. On the experience with guideposts in the sixties, see J. Sheahan, *The Wage-Price Guideposts*, Brookings Institution, 1968; the various papers in G. P. Schultz and R. Z. Aliber, eds., *Guidelines, Informal Controls and the Market Place*, University of Chicago Press, 1966; G. L. Perry, "Wages and the Guideposts," in *American Economic Review*, September 1967, pp. 897–904, and M. Bronfenbrenner, "Guidepost-Mortem," in *Industrial and Labor Relations Review*, July 1967, pp. 637–49. For a view of incomes policy over the period since World War II, see A. R. Braun, "The Role of Incomes Policy in Industrial Countries Since World War II," in *International Monetary Fund Staff Papers*, March 1975, pp. 343–83, and C. D. Goodwin, editor, *Exhortation and Controls: The Search for a Wage-Price Policy, 1945–1971*, The Brookings Institution, 1975.

[7]*Annual Report*, January 1967, p. 128, and repeated verbatim in *Annual Report*, February 1968, p. 126. See also *Annual Report*, January 1969, pp. 118–21.

[8]From statement before the Senate Committee on Banking, Housing, and Urban Affairs, February 7, 1973, reprinted in *Federal Reserve Bulletin*, February 1973, p. 81.

FISCAL POLICY

The deliberate use of fiscal policy as a possible means of attaining and maintaining full employment and a stable price level is a development of the past four decades. This use of fiscal policy began during the thirties, largely as a result of three developments: the apparent ineffectiveness of monetary policy as a means of overcoming the severe unemployment of the Great Depression, the "new economics" advanced by Keynes with its emphasis on aggregate demand, and the growing importance of government spending and taxation in relation to the economy's total income and output. From its relatively modest beginnings, fiscal policy has grown to be a major means by which the government attempts to achieve high employment and to prevent inflation. As noted at the beginning of the chapter, the Employment Act of 1946 directed legislatively that fiscal policy be used toward the achievement of these ends.

The success of the Keynesian economics was such that from the forties into the sixties there was little question but that government could raise or lower aggregate demand through appropriate changes deliberately brought about in government purchases, transfers, and tax collections. Starting in the sixties, monetarism, a doctrine that will be outlined in the following chapter, achieved a measure of success in its general attack on Keynesian economics. As part of the attack, the monetarists brought into question the basic tenet of Keynesian economics which had long gone unquestioned. They revived an old classical notion that holds that, except under certain conditions, increases in government spending do not add to total spending but simply supplant or "crowd out" an equal amount of private spending. A parallel argument is made for decreases in tax rates. To the extent that such crowding-out occurs on anything like a dollar-for-dollar basis, fiscal policy becomes powerless to affect aggregate demand. All the other problems encountered in trying to use fiscal policy to deliberately influence aggregate demand then become academic or of no practical importance.

Most economists, however, do not accept the crowding-out argument in anything like the form advanced by some monetarists; in other words, most economists accept that fiscal policy can be used to vary aggregate demand in a way and to a degree that contributes to economic stabilization. If it can indeed be so used, it then becomes necessary to face the many real-world problems that complicate the planning and execution of actual fiscal policies. A problem that, in a sense, underlies all other such problems is that of evaluating the impact of any overall fiscal program on the level of economic activity. To measure in a meaningful way the stimulating or restraining influence of any actual federal fiscal program or of the federal budget as a whole for any time period requires recourse to what is called the *full-employment budget surplus*. The first part of this section is devoted to an examination of this concept.

The full-employment budget surplus gives us a measure of the stimulus or restraint exerted by a particular fiscal program, but suppose that our objective is to provide more or less stimulus or more or less restraint than that indicated by a given fiscal program. The practical problem then faced is whether the program can be altered fairly promptly in a way that will yield the desired result. This is essentially the problem of flexibility in fiscal policy, and is the subject of the second part of this section.

Any adjustment in the fiscal program to vary the restraint or stimulus exerted by that program calls for changes in the level and perhaps the composition of government purchases, transfer payments, or taxes or in various combi-

nations of these. The practical difficulties involved in varying expenditures and taxes in the way that may be required if fiscal policy is to contribute to the stabilization of the economy are briefly considered in the last part.

The Full-Employment Budget Surplus

On the basis of the elementary fiscal models presented in Chapter 6, we reached the straightforward conclusion that fiscal changes that involve a deficit are expansionary and fiscal changes that involve a surplus are contractionary. If one works with this kind of model, it would seem that all he need do to determine whether the impact of the government budget is expansionary or contractionary in any period is to note whether it shows a deficit or a surplus for that period. If it shows a deficit but a smaller one than in the preceding period, the rule suggests that the budget is still expansionary but less so than it was in the preceding period. Similarly for other period-to-period changes in the size of the surplus or for changes from surplus to deficit and deficit to surplus.

In these pages we will examine a major qualification to this rule, one that shows, under certain circumstances, that a rise in the deficit from one period to the next is not indicative of a more expansionary budget, as suggested by the rule, but of the very opposite.[9]

The federal government has a budget program for each year that fixes both planned expenditures and tax rates. This program cannot, however, fix in advance but can only estimate the size of the deficit or surplus, because that will depend in part on the level of economic activity, which is not known in advance. Given the possibility that economic activity may vary over a sizable range, an unchanged program of planned government expenditures and tax rates is accordingly consistent with a whole range of possible surpluses or deficits. The nature of this relationship is illustrated with hypothetical figures in Figure 24-1.

The horizontal axis of this figure shows actual GNP as a percentage of full-employment GNP, which is the estimate of the level of GNP that would be realized if the economy were operating at full employment. The vertical axis expresses the federal surplus or deficit as a percentage of full-employment GNP. Each of the two lines shown describes a different program of planned government expenditures and tax rates, the upper line involving a smaller amount of planned expenditures and/or higher tax rates than the lower line. The upper line shows that the particular expenditure and tax program that determines the position of that line will result in a deficit equal to 0.5 percent of full-employment GNP if the actual level of economic activity is at 96 percent of the full-employment level, and a surplus equal to 0.5 percent of full-employment GNP if the actual level of economic activity is at 98 percent of full employment. These two situations are shown as *C* and *A*, respectively.

Now let us assume that the economy happens to be operating at the 98-percent level in

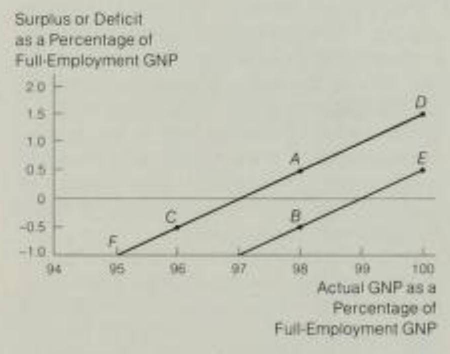


FIGURE 24-1
The level of economic activity and the federal surplus or deficit

[9]A review of the third fiscal model presented in Chapter 6 will contribute to an understanding of the material that follows.

the current period and that it is expected to remain at this level in the following period (assuming the absence of monetary and fiscal policy actions taken to move it to a higher level). In this event, with a budget program that gives us the upper line, a surplus equal to 0.5 percent of full-employment GNP is indicated at *A*.

Suppose, however, that the income level turns out to be only 96 rather than the 98 percent of the full-employment level that was expected. With an unchanged budget program, the budget has shifted from the surplus at *A* to the deficit at *C*. In this case, what must be seen is that the deficit is the result solely of unplanned, nondiscretionary, or automatic changes in tax receipts and expenditures. The deficit is passively induced by the slowdown in economic activity. Graphically what we find is simply a movement back down the upper line that designates the given unchanged budget program.

Consider next the opposite case of a discretionary change in the budget program, or a change that involves a shift in the line. Again we begin with the economy operating at the 98-percent level and with the budget program shown by the upper line. As before, this results in the surplus at *A*. Now, however, we assume a discretionary change in the budget program, such as a cut in tax rates or a planned increase in government expenditures. The height of the line reflects the basic budget program, and changes such as those noted would involve a downward shift of the line, say, to the position of the lower line. If the income level remained at 98 percent of the full-employment level, this new budget program would result in the deficit at *B* instead of the surplus at *A*.

Note that the deficit at *C* is equal to the deficit at *B*, but they nonetheless differ completely in what they say about the impact of the budget program. The movement from the surplus at *A* to the deficit at *C* is the result solely of the automatic decrease in tax receipts and the unplanned increase in expenditures for unemployment compensation that accompany the slowdown in economic activity. A deficit that originates in this way does not indicate any change in the impact of the existing budget program. On the other hand, the movement from the surplus at *A* to the deficit at *B* is the result solely of discretionary changes in the budget program that make the impact of that program more expansionary or less contractionary than it previously was.

In practice, a given change in the actual deficit or surplus from period to period may be the result of changes in both the budget program and the level of economic activity. Furthermore, a change in the level of economic activity may itself be the result of a change in the budget program—that is, a change in government spending—that is not offset by an opposite change in private spending. For a simple illustration, begin with the economy operating at the 98-percent level with a deficit at *B*, and assume a change in the budget program that shifts the line to the position of the upper line. If there were no change in the level of economic activity, there would be a movement from the deficit at *B* to the surplus at *A*, but the very increase in tax rates or decrease in expenditures that underlie the shift in the line are likely to cause a change in the level of economic activity.

Suppose the change in the budget program that produces this upward shift of the line were sufficiently restrictive, other things being given, to cause economic activity to fall all the way to the 95-percent level at which we find the deficit at *F*. In this event, we have a movement from the deficit at *B* not to the surplus at *A* but to the even larger deficit at *F*. If the impact of the budget is judged solely in terms of the size of the deficit or surplus, we would conclude incorrectly that we have a less restrictive or more expansionary budget program. On the other hand, if the impact of the budget is judged in terms of all the relevant information, we would conclude correctly that there has been a shift toward a more restrictive or less expansionary budget program. With all the relevant information at hand, we see in the present case that

the rise in the deficit indicates not a more expansionary or less restrictive budget program but actually the very opposite of this.

This illustration resorts to an extreme case to emphasize the point, but the change in the size of the deficit or surplus is not a reliable guide to the change in the impact of the budget in the general case as well. One way to a more reliable guide is to remove the influence of changing levels of economic activity on the deficit or surplus consistent with any given budget program. In so doing we derive what is called the full-employment budget surplus or deficit, a measure of the surplus or deficit that reflects only changes in the budget program and therefore indicates changes in the contractionary or expansionary impact of the budget program itself. This particular measure of budget surplus and deficit was introduced by the Council of Economic Advisers in 1962 and has been developed over the years since then.[10]

In formal terms, the full-employment budget surplus may be defined as the federal budget surplus, on a national income accounts basis, that would be generated by a given budget program if the economy were operating at full employment with stable prices throughout the year. Otherwise expressed, it is the difference between federal receipts and expenditures calculated for existing expenditure programs and tax rates, but with expenditures for unemployment compensation adjusted to what they would be at full employment and with government receipts under existing tax rates adjusted to what they would be at full employment. Figure 24-1 shows at *D* and *E* the full-employment budget surplus (expressed there as a percentage of full-employment GNP) that follows from the two budget programs represented by the two lines in that figure.

To estimate the size of the full-employment budget surplus or deficit for any quarter or year clearly calls for a series of other estimates of a kind suggested by the second definition above. First, we must have an estimate of what the GNP would be at full employment. The basic method of arriving at this estimate is the growth rate extrapolation method, which is a simplification of what is popularly known as Okun's law.[11] This method assumes that full-employment GNP in real terms grows at a fairly constant rate over extended periods of time. Once the rate is determined and a base year in which there was full employment is selected, the real full-employment GNP for other years is calculated from a trend line of production consistent with an average unemployment rate of 4 percent of the civilian labor force, the 4-percent rate being taken as the rate consistent with full employment.

In 1976 the Council of Economic Advisers introduced some major modifications to this procedure in producing its estimates. It had previously been assumed that full utilization of other resources like land and capital would accompany a 4-percent unemployment rate. The new procedure for estimating full-employment output allows explicitly for the contribution of fixed capital. Full employment of fixed capital

[10]The concept actually originated in the mid-forties, but it received emphasis by the council starting in 1962. (See the *Annual Report of the Council of Economic Advisers*, Jan. 1962, pp. 78–84.) A more thorough discussion is found in M. E. Levy, *Fiscal Policy, Cycles and Growth*, National Industrial Conference Board, 1963, Ch. 6, and A. M. Okun and N. H. Teeters, "The Full Employment Surplus Revisited," in *Brookings Papers on Economic Activity*, 1, 1970, pp. 77–110. This article, on pp. 84–102, provides a good discussion of some of the major problems associated with the full-employment surplus. See also A. S. Blinder, *Fiscal Policy in Theory and Practice*, General Learning Press, 1973, pp. 6–12. For more on the estimating procedure that is touched on below, see N. H. Teeters, "Estimates of the Full Employment Surplus, 1955–64," in *Review of Economics and Statistics*, Aug. 1965, pp. 309–21; and K. Carlson, "Estimates of the High-Employment Budget, 1947–67," in *Review*, Federal Reserve Bank of St. Louis, June 1967, pp. 6–13.

[11]Okun's law relates total output to labor-force utilization and productivity. For an analysis of this relationship, see A. M. Okun, "Potential GNP: Its Measurement and Significance," in *Papers and Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 1962, pp. 98–104. See also L. C. Thurow and L. D. Taylor, "The Interaction Between the Actual and the Potential Rates of Growth," in *Review of Economics and Statistics*, Nov. 1966, pp. 351–60.

is assumed to exist when the utilization index of manufacturing capacity prepared by the Department of Commerce is 86 percent. In addition to introducing the capacity utilization rate into the calculation, the Council in 1976 also replaced the earlier fixed 4-percent figure as a definition of full employment of the labor force with a figure that varies over time. This is to adjust for such changes over recent years as the increase in the fraction of the labor force made up of women and younger workers, groups which are subject to higher than average unemployment rates. The estimate is that these changes raise the full-employment unemployment rate to 4.9 percent in 1976. Potential GNP in 1976 is then the output that would be produced if the unemployment rate were 4.9 percent and the capacity utilization rate in manufacturing were 86 percent. For the 1962–76 period, the new procedure reduces the average annual growth rate of full-employment output to 3.6 percent from the 3.9 percent it would be under the other procedure. For 1976 in particular it yields a figure for full-employment GNP that is about 4 percent less or in 1972 prices $58 billion less than the estimate of potential for that year under the other procedure.[12]

The estimate of full-employment GNP for each year yielded by this procedure is in constant dollars. The full-employment budget surplus or deficit for each year is to be expressed in current dollars, so the estimates of full-employment GNP are converted into current dollar estimates by multiplying by the GNP implicit price deflator. Given the estimate for full-employment GNP in current dollars, the second step is to estimate the amount of revenue that would be generated at this level of GNP under the existing tax structure and tax rates. This requires estimates of such variables as the proportions of full-employment GNP going to corporate profits and to personal income and then the application of an appropriate tax rate to each income component. As an example of one of the many difficulties that appear at this point, the appropriate tax rate under the personal income tax depends on the size of the tax base to which it is applied, and the size of this base requires an estimate of how the distribution of income between personal income and corporate profits will change at full employment from what it is at a position below full employment. Last and least troublesome in this series of estimates is the estimation of government expenditures at full employment. This reduces to an estimate of what the amount of unemployment-compensation payments will be at the full-employment level of GNP, since other budget items are determined by administration proposals and congressional action and, with stable prices, may reasonably be treated as independent of the GNP level.

Table 24-1 shows the actual receipts, expenditures, and deficit and the Council's estimates of full-employment receipts, expenditures, and deficit or surplus for the federal government for calender years 1970–76. The actual figures are based on the national income accounts budget. During all of these years the actual budget was in deficit.[13] With the single exception of 1974, the same is true for the full-employment budget. While it is apparent that deficits predominate in these years, what is significant is the difference between the deficits in the two budgets for each year. The year 1970 was a year of recession in which the actual deficit was $12.1 billion, but the full-employment deficit was only $2.6 billion. Therefore, according to the Council's estimate, the budget was stimulative in 1970, but the amount of stimulus is indicated not by the actual deficit of $12.6 billion but by the full-employment deficit of $2.6 billion. Unlike the figure for the full-employment budget deficit, whose very computation has re-

[12] See *Annual Report of the Council of Economic Advisers*, January 1977, pp. 48–55.

[13] Since 1958 there have been only four calendar years (1960, 1963, 1965, and 1969) in which the actual budget showed a surplus, and the cumulative surplus for those four years was $12.3 billion, little more than a fifth of the deficit for 1976 alone.

moved the effect of that year's departure from full employment of the labor force and fixed capital, the figure for the actual budget deficit reflects the effect of the cyclical downturn and, for reasons discussed earlier, thereby obscures the extent to which the budget program itself is stimulative or restrictive. Another measure of the stimulus or restriction that results from the budget is given by the *change* in the full-employment deficit or surplus. From 1969 to 1970, the change of −$6.3 billion indicates that whatever its impact in 1969 (not shown here), the budget in 1970 shifted in the direction of stimulation by the amount of $6.3 billion.

For a more dramatic illustration consider the year 1974, the most severe year of contraction in post-World War II experience. The actual budget showed a deficit of $11.5 billion, but the full-employment budget showed a surplus of $14.1 billion. Hence, the conclusion is that in that year of sharp recession the federal budget did not exert a contracyclical effect but just the opposite. From a full-employment deficit of $7.9 billion in 1973 to a full-employment surplus of $14.1 billion in 1974 was a one year budget swing of $22 billion toward contraction when what was needed was a swing toward expansion. As a result of a tax rebate and other discretionary fiscal actions in 1975, the full-employment surplus of $14.1 billion of 1974 was replaced with a full-employment deficit of $12.5 billion to provide a $26.5 billion swing in the expansionary direction between those two years, a swing that contributed to the economic recovery that got under way in the spring of 1975.

Apart from errors in estimating the full-employment budget surplus or deficit, we may accept what these illustrations are intended to show: as a general rule, the size of a full-employment budget surplus or deficit is, in and of itself, an indicator of how restrictive or stimulative a budget program is during that period; and the period-to-period change in this magnitude is indicative of the changing thrust of the budget program between those periods. While this much is acceptable, one cannot go on from this to accept that the goal of fiscal policy

TABLE 24-1
Actual and full-employment federal government receipts, expenditures, and deficit or surplus, calendar years, 1970–76
(billions of dollars)

	1970	1971	1972	1973	1974	1975	1976[a]
Actual:							
Receipts	192.1	198.6	227.5	258.3	288.2	286.5	330.6
Expenditures	204.2	220.6	244.7	265.0	299.7	357.8	388.9
Surplus or deficit (−)							
Amount	−12.1	−22.0	−17.3	−6.7	−11.5	−71.2	−58.3
Change	−20.6	−9.9	4.7	10.6	−4.8	−59.7	12.9
Full Employment:							
Receipts	201.0	210.0	222.1	257.5	311.8	337.6	371.6
Expenditures	203.6	219.1	243.6	265.4	297.7	350.1	381.9
Surplus or deficit (−)							
Amount	−2.6	−9.2	−21.5	−7.9	14.1	−12.5	−10.3
Change	−6.3	−6.6	−12.3	13.6	22.0	−26.5	2.2

[a]Preliminary figures
SOURCE: *Annual Report of the Council of Economic Advisers*, January 1977, Table 18, p. 76.

should be neither a full-employment budget surplus nor a deficit but a budget program that is, in a sense, neutral. However, during the Nixon administration, this was often stated to be the goal. The traditional objective of an annually balanced budget was replaced with the objective of a budget balanced at full employment. This is stated in the president's budget message for the fiscal year 1973 in the following words: "The full-employment budget concept is central to the budget policy of this Administration. Except in emergency conditions, expenditures should not exceed the level at which the budget would be balanced under conditions of full employment. The 1973 Budget conforms to this guideline. By doing so, it provides necessary stimulus to expansion, but is not inflationary . . . ."[14] The same idea was conveyed in the budget message for 1974: "I am proposing to avoid both higher taxes and inflation by holding spending in 1974 and '75 to no more than revenues would be at full employment."[15]

What the administration set as its objective was actually a budget that was neither stimulative nor restrictive, but this objective can turn out to be a rejection of the use of the budget as a means of attaining and maintaining full employment and preventing inflation. That requires a budget that is at times stimulative and at times restrictive or a full-employment budget that at times shows a deficit and at times shows a surplus. For example, suppose that planned private saving is expected to be greater than planned private investment at the full-employment level of output. Then a deficit in the full-employment budget is required to provide equilibrium at full employment. In the opposite case, if planned private investment is expected to be greater than planned private saving at full employment, a surplus in the full-employment budget is required to prevent inflation. Conditions may be such that a balance in the full-employment budget is appropriate policy, but the opposite is more likely to be true as private investment and saving plans do not automatically balance at the full-employment level of output.

If one grants that the federal budget is to be used in a way that contributes to economic stabilization, the concept of the full-employment budget surplus appears essential in planning appropriate fiscal actions designed to offset an excess or deficiency of planned private investment relative to planned private saving at the full-employment level. Such fiscal actions should thereby help the economy to reach full employment if it is not already there or, if it is already there, to prevent inflation. There are, of course, serious problems in using the concept. The very calculation of the full-employment budget surplus is itself fairly crude, and small changes in this surplus or deficit are probably not significant as more than indicators of whether the fiscal program is moving toward more or less restraint or stimulus from one quarter to the next. A more serious problem, found in an economy below full employment, is that of estimating what private saving and private investment would be at full employment in order to know what the appropriate full-employment budget surplus or deficit should be. Still another difficulty is that of securing the necessary discretionary flexibility in expenditures and/or taxes, a question we will turn to below.

Before we do, let us consider briefly a concept closely related to that of the full-employment budget surplus—the concept of *fiscal drag.* Underlying this notion is the fact that federal revenue rises more than proportionally with increases in GNP, especially for short-run increases in GNP, due mainly to the progressive nature of the personal income tax and the high responsiveness of corporate profits to changes in GNP. If federal expenditures increase at the same rate as GNP, the result is a budget program whose revenue side increases more than

[14]See *U.S. Budget in Brief,* Fiscal Year 1973, U.S. Government Printing Office, p. 9.

[15]See *The Budget of the United States Government,* Fiscal Year 1974, U.S. Government Printing Office, p. 21.

its expenditure side and a budget that therefore exerts a drag on the growth of GNP. If the economy is already at full employment, a budget program such as this will show a smaller deficit or larger surplus each year. Since there is no reason to expect that this will be just offset by an equal change in the difference between private saving and private investment, this shrinking deficit or expanding surplus will eventually force the economy below full employment. If it already is below full employment, this same budget program will prevent the attainment of full employment. The best illustration of the latter situation was found in the federal expenditure program and tax structure as they were before 1964. An expansion of GNP toward the full-employment level would automatically increase tax receipts by more than the increase in expenditures and thus enlarge a full-employment budget surplus which existed then. As we have seen, unless there was to be a corresponding excess of planned private investment over planned private saving at the full-employment level to offset such a full-employment budget surplus, this surplus itself would prevent the attainment of full employment. It is in this sense a "drag" on the very expansion of output and employment.

The full-employment budget was in fact highly restrictive during the years preceding 1964, the year in which Congress approved a tax cut that the administration had called for in 1962. The reduction in taxes, combined with the expenditure program as it was then, wiped out the full-employment budget surplus in 1965 and also, given the magnitudes of private saving and investment, enabled the economy to achieve a full-employment level of output in 1966 for the first time since 1953. From the perspective of the late seventies, the prospect that the attainment of full employment may again be blocked by fiscal drag appears extremely remote for the years just ahead. Tax rates may not be reduced to offset the expansion of tax receipts that occurs as the economy moves closer to full employment, but the Congress will have little difficulty in matching those increased tax receipts with increased expenditures. The demand for federal government spending for education, housing, pollution control, military procurement, job programs, and for grants to state and local governments to meet welfare, crime control and other local services is such that there is little question that federal expenditures will rise as rapidly as federal revenues. Given the attitude of the majority of the U.S. Congress toward expanding expenditures, for the near future at least the earlier problem of "fiscal drag" has been replaced by the opposite problem of "fiscal squeeze"—not too much tax receipts relative to expenditures but too much expenditures relative to tax receipts.

Still, in the longer view it is possible that the problem of fiscal drag may return and the challenge to full employment that it presents will again have to be faced. However unlikely this now seems, should we somehow some day reach a situation of this kind, the federal government will then have to either continuously expand its expenditures as a fraction of GNP or continuously cut tax rates or both. To do none of these things in such a situation would be to permit once again the occurrence of recession and unemployment that follows from fiscal drag.

Flexibility of Fiscal Policy

Planned private saving may exceed private investment at full employment so that a full-employment budget deficit is required if full employment is to be realized. If the actual situation happens to be one of a full-employment budget surplus and thus the actual level of income is one below full employment, the need seems clear: discretionary changes in government expenditures and/or tax rates of the order required to produce the full-employment budget deficit that is consistent with the full-employment level of income. Meeting the need

is another matter. Here we look briefly into the particular question of whether we can secure the required degree of flexibility in government expenditures and/or tax rates to produce now a deficit and then a surplus as may be required to meet the needs of the situation. This same question will arise in the following chapter in connection with monetary policy.

If we start off with a comparison of the relative flexibility of fiscal and monetary policy, it is generally conceded that monetary policy has the advantage here. Although this is due in part to the inherent flexibility of certain tools of monetary management, it is also due in part to the fact that the decision-making authority lies in an essentially independent agency whose political aims are limited to its own perpetuation and the preservation of its customary role and whose power is concentrated in the hands of a few men. It is the seven-man Board of Governors of the Federal Reserve System or the twelve-man Federal Open Market Committee (which includes the same seven men plus the presidents of five of the Federal Reserve Banks) that makes the decisions on monetary policy in the United States. The presidents of the other Federal Reserve Banks, the Secretary of the Treasury, the Council of Economic Advisers, and others influence these decisions, but the power is nonetheless concentrated in these few hands. Although Congress in 1975 adopted a resolution that restricts somewhat the complete discretionary power previously possessed by these men, they still have the power to alter policy promptly and over a wide range. Their performance record is, of course, another matter; here we simply note the flexibility with which discretionary changes can be made in the area of monetary policy.

In contrast, the decision-making process in the area of fiscal policy involves, in a sense, the whole of the executive and legislative branches of the federal government. The political motivation here is also obviously of an altogether different kind, and decisions in the fiscal policy area, especially decisions involving higher tax rates, are not made without allowance for their effect on the electorate in the election that follows. Apart from this political bias, which is perhaps unavoidable in a system like ours, a certain lack of maneuverability follows simply from the fact that the decision-making power in the fiscal area does not rest in the hands of a small group. The President has some limited power through his ability to control, to a degree, the timing of federal fiscal actions by speeding up or delaying expenditures and tax refunds. However, to get flexibility of a degree at all comparable with that found in the area of monetary policy would require that the President or some small group be given limited but discretionary power to vary certain tax rates and perhaps also some expenditure programs. Presidents Kennedy and Johnson both made proposals along this line. In his final budget message in January 1969, President Johnson suggested that "consideration should be given to establishing as a permanent part of our tax system an element of flexibility under which the President . . . subject to congressional veto, would have discretion to raise or lower personal and corporate income-tax rates within specified limits—such as 5 percent in either direction."[16]

Congress traditionally has closely guarded its power over the tax structure, tax rates, and expenditure programs, and it did not act favorably on either the Johnson proposal or a very similar proposal made earlier by President Kennedy. And despite the fact that the election of 1976 put a Democrat in the White House for the first time since 1968, it is doubtful that the heavily Democratic majority in the Congress would react favorably to a proposal of this kind from President Carter, no matter how well the Congress might get along with him on other matters. In any event, it is clear that discretionary fiscal policy will remain less flexible than it might be as long as Congress retains almost complete power to make significant discretionary changes. It will then continue to depend on legislative action, and legislative action, at

[16] *The Budget of the United States Government*, Fiscal Year 1970, U.S. Government Printing Office, 1969, p. 12.

least that which calls for the politically unpopular decision to raise tax rates or cut back on various spending programs, will probably continue to be preceded by time-consuming congressional hearings and debate.

Under these conditions, it is the built-in or automatic stabilizers on which we will primarily depend for whatever real, short-run flexibility we have in fiscal policy. It is these, we will see, that underlie the discussion in the preceding section of surpluses and deficits that automatically arise as the level of income increases and decreases. Beyond automatic flexibility is formula flexibility, a type of flexibility that would not require Congress to give the President the amount of power he would acquire with even limited discretionary authority. In the following pages we will take a brief look at the way fiscal policy now operates, with some flexibility provided by built-in stabilizers, and consider how this flexibility might be increased through the formula approach.

Built-in Flexibility Built-in flexibility is achieved when changes in tax collections and government spending vary automatically, promptly, and in the right direction to produce a stabilizing effect on aggregate demand. Automaticity means that no specific action need be taken; promptness means that there is little lag between changes in aggregate demand and changes in government spending and tax collections; right direction means that decreases in aggregate demand call forth additional government spending and reduced tax collections, while increases in aggregate demand call forth the opposite. In general, when aggregate demand and income are rising, automatic and prompt increases in tax receipts and decreases in transfer payments tend to dampen the expansion; when aggregate demand and income are falling, automatic and prompt decreases in tax receipts and increases in transfer payments tend to dampen the contraction.

These tendencies follow from the way in which total tax collections and government transfer payments automatically vary to prevent disposable personal income from rising or falling as much as GNP. In the case of a recession and falling GNP, for example, disposable personal income is protected somewhat because tax collections automatically fall and transfer payments automatically rise. Disposable personal income, therefore, falls less than it would if these stabilizers were not in operation. With disposable personal income so protected, personal consumption expenditures will fall less than they would if these stabilizers were not in operation. Finally, by avoiding what would otherwise be a sharper decline in consumption expenditures, the cumulative fall in GNP itself is less than it would be if these automatic stabilizers were not in operation. In simplest terms, this is the way built-in stabilizers produce a smaller fluctuation in GNP than would be the case in their absence.

This dampening of income movements is the usual consequence of the operation of built-in stabilizers, but it is also possible for some to turn into built-in destabilizers and operate perversely or accentuate income movements. Such was the case in 1974, the year of an extraordinary combination of a sharp decline in real GNP and a "double-digit" rate of inflation. The dollar volume of transfer payments in the form of unemployment compensation benefits responded in a stabilizing manner, because this amount varies directly with the number of eligible unemployed workers, a number that increased with the real decrease in real GNP. However, income-tax collections responded in a perverse or destabilizing manner. From 1973 to 1974, real GNP decreased 1.7 percent while nominal GNP increased 8.2 percent. Many taxpayers found that their nominal incomes increased substantially at the same time that their real incomes decreased, and it is nominal income on which their income taxes are based. Therefore, instead of the tax system taking a smaller fraction of such smaller real incomes, the result needed for the tax system to act in a stabilizing way, it took a larger share of such smaller real incomes, a result that shows the tax system to act in a destabilizing way. Some

previous recessions had also been accompanied by inflation, but none to anything like the degree found in 1974. The perverse effect of the tax system was truly vicious in that year.[17]

Even when such a perverse effect is absent and the built-in stabilizers live up to their name, the dampening effect they can exert on income movements is not an unmixed blessing. Whereas the resistance the built-in stabilizers provide to a downward movement is desirable in an underemployed economy, the resistance they provide to an upward movement in the same economy is undesirable. This, as we saw earlier, is termed fiscal drag and calls for appropriate offsetting actions in the form of discretionary changes in government expenditures or tax rates. In an economy at full employment, the built-in stabilizers would tend to have a stabilizing effect in both directions. They would offer resistance to the worsening of any downturn brought on by a sharp decline in demand and to a worsening of any inflationary movement brought on by a sharp expansion in demand. In an economy operating at full employment, they provide a limited but helpful buffer against cumulative movements in either direction.

Economists quite generally favor the greatest possible use and strengthening of built-in stabilizers where possible. If one could ignore other public policy objectives, any number of changes could be adopted that would strengthen existing stabilizers. In the case of the personal income tax, for example, a reduction in the size of the personal exemption and a tightening of deduction provisions would put a larger part of personal income into the base on which personal income taxes are computed, thereby increasing the cyclical variability of the revenue from this tax. A more progressive rate structure might contribute somewhat to the same end. In the case of indirect taxes, an *ad valorem* tax would increase the effectiveness of these stabilizers, since the revenue yield would vary more over the cycle under *ad valorem* than under *specific* taxes. In the case of government transfer payments, unemployment compensation payments could be made a more effective stabilizer by enlarging weekly benefits, lengthening the period over which these benefits may be received, and increasing the number of workers covered, actions that were taken during the especially severe recession that ended in 1975. Although each of these and a number of other changes can be made for the purpose of improving the effectiveness of an automatic program of stabilization, many such changes are ruled out because they conflict with and are judged subordinate to other aims of public policy.[18]

[17]Apart from the fact that a major built-in stabilizer becomes perverse under these conditions, elementary tax justice dictates that steps be taken to avoid such a result. Essentially all that is required is that the size of the personal exemption (or the size of the personal credit, if the exemption is replaced with a credit as proposed by President Carter), the size of the standard deduction, and the width of the income tax brackets be adjusted upward periodically in line with the increase in an appropriate price index. Taxpayers would then move into higher tax brackets as they earned higher real incomes, not as they earn higher nominal income as is now the case. Canada adopted such an indexation system, but the U.S. Congress has displayed little interest, the only apparent reason being that it is to the selfish interest of the Congress to maintain a system that gives the government the additional revenue that comes with what amounts to higher tax rates without facing the political repercussions that would be faced if the Congress were to openly raise tax rates.

[18]For an analysis of the quantitative impact of specific built-in stabilizers over a series of business cycles, see M. O. Clements, "The Quantitative Impact of Automatic Stabilizers," in *Review of Economics and Statistics*, Feb. 1960, pp. 56–61. An analysis of the actual behavior of each of the major built-in stabilizers will be found in the second chapter of W. Lewis, Jr., *Federal Fiscal Policy in the Postwar Recessions*, Brookings Institution, 1962. See also P. Eilbott, "The Effectiveness of Automatic Stabilizers," in *American Economic Review*, June 1966, pp. 450–65. For a study of the amount of built-in flexibility provided on the side of federal government spending, see N. H. Teeters, "Built-in Flexibility of Federal Expenditures," in *Brookings Papers on Economic Activity*, 3, 1971, pp. 615–48. For example, the Medicaid program with its close connection to the welfare rolls works as a built-in stabilizer. Another stabilizer results from the growth in the number of older people eligible for benefits under retirement programs. When jobs are lost during recession, for many older persons retirement is frequently more socially acceptable than unemployment.

Partly because there are limits to how much automatic stabilization can be built into the system without causing conflict with other aims of public policy, most economists and many business persons today are unwilling to limit the use of stabilizing fiscal policy to what can be accomplished through its passive role in an automatic program alone.[19] Instead, they generally favor an active role in which taxes and government spending (purchases of goods and services as well as transfer payments) are made to vary according to formula or executive discretionary action. The argument for a more active role for fiscal policy is substantially strengthened when we take into account a major limitation of the most complete system of built-in stabilizers—namely, that the stabilizers in themselves cannot prevent a downturn from occurring because they do not come into effect until there already is some downturn in spending and income. True, they can help prevent a downturn from growing cumulatively worse, but they cannot *in themselves* reverse a downturn and initiate an expansion. On the other hand, although subject to other limitations, nonautomatic programs in which expansionary fiscal measures are initiated in the expectation of a downturn may in themselves (assuming correct forecasting) prevent downturns. Formula flexibility is like the built-in stabilizers in the sense that it is essentially automatic. However, it may be designed to allow a greater scope for stabilizing action than is provided by the built-in stabilizers, and, unlike the built-in stabilizers, may be able to reverse a downturn once begun.

[19]The most influential group that long favored the restriction of stabilizing fiscal measures to those of a purely automatic nature is the Committee for Economic Development. The CED took this position in its first statement on the subject in 1947 in *Taxes and the Budget: A Program for Prosperity in a Free Economy*. Through a long series of national policy statements, the CED by 1958 had recognized the need for some discretionary action in its *Anti-Recession Policy for 1958*. Still, fiscal measures were assigned the inherently passive role provided by built-in stabilizers, the active role to be played by monetary measures. In a 1969 statement called *Fiscal and Monetary Policies for Steady Economic Growth*, the CED took a large step toward discretionary fiscal policy and toward providing flexibility in this policy by recommending that Congress give the President power to vary personal and corporate income tax rates by up to 10 percent. With a few additional safeguards, the CED proposal was similar to that made by President Johnson in his final budget message.

Formula Flexibility Formula flexibility relies on changes in selected indexes such as the unemployment rate or the consumer price level as indicators of a need for specific changes in income tax rates, transfer payments, or even public works expenditures. Formula flexibility is like built-in flexibility in that action takes place automatically in response to realized changes in the business situation, but the two differ in that built-in flexibility is obtained within the existing tax and transfer payments structure while formula flexibility, once activated, changes the structure itself. For example, in oversimplified form, the formula could require a specified reduction in personal income tax rates when the unemployment percentage equaled or exceeded 5 percent for two consecutive months, or it might call for an increase in those tax rates when the consumer price index rose 5 percent in a specified time interval. Similarly, a formula could call for liberalizing unemployment compensation and other kinds of government transfer payments in response to increases in the unemployment rate. A number of small-scale public works projects could also be kept "on the shelf" until there were specific changes in the unemployment rate or in other indexes of business conditions. Apart from public works with their focus on construction, emergency employment programs that hire extra people for things like park, recreation, fire protection, and security services may similarly be triggered in the same way. For example, a 1971 act provided for a public employment program that would take effect when the national unemployment rate was 4.5 percent or higher for a three-month period. Of a different nature, a 1974 act provided unemployment benefits during 1975 for some workers ineligible for the regular state or federal programs in the event that the local

area unemployment rate averaged 6.5 percent or more for three consecutive months or the national unemployment rate averaged 6.0 percent or more for the same time period. By relying on a formula, changes in tax rates and expenditure programs would occur with minimum administrative delay and without forecasts of future business conditions—forecasts that, as is well known, are often wrong.

Yet, despite the automaticity of action associated with rigid adherence to a formula, the action so triggered may be wrong just as often as action that is based on forecasts. If an unemployment rate that is above some specified level for a period of a few months is the signal for action, for example, corrective action may come just at the time when the forces that caused the high unemployment are dying out. There is no way of knowing such things with certainty in advance, but detailed study of the situation may provide, at the minimum, some understanding of the cause or causes of the business downturn and the high unemployment and hence some clue as to whether a turnabout may be near at hand. For example, if study suggests that the downturn resulted primarily from an inventory liquidation, an end to the downturn may not be many months off. If, on the other hand, study shows that decreasing plant and equipment expenditures are the cause, the situation may be more serious, and the action indicated by the formula or even more vigorous action may be in order. Without in any way detracting from the advantages of the formula device, action triggered by formula alone, without study of the total business situation, would receive the support of few economists. Actually, some economists support formula flexibility more for political than for economic reasons. They would favor granting the President limited authority to take discretionary action, but realize that Congress is not likely to delegate such authority.

In sum, economists usually conclude that the role of fiscal policy in limiting short-run fluctuations in economic activity is for the most part restricted to the automatic or built-in stabilizers. Though action is taken from time to time by Congress itself, such action has typically not been sufficiently flexible to be useful against short-run fluctuations in economic activity. Although this was not strictly an anti-cyclical action, a classic illustration of belated action is the 1964 tax cut. That cut requested by President Kennedy in 1962 to raise the economy's growth rate did not occur until after almost two years of deliberation by Congress. Another illustration is the experience of a few years later when the reverse kind of action was needed to meet the threat of inflation following from the large rise in military expenditures for Vietnam. It again took about two years for action to take place. However, the last few years may have brought a change. Congress seems to be able to take discretionary action much more quickly than heretofore, at least when the kind of action called for is to cut taxes or increase spending. Thus, once the seriousness of the 1974 downturn was recognized, Congress in fairly quick order enacted a tax cut bill in March of 1975 which reduced that year's tax collections about 5 percent below what they otherwise would have been. It may be that there is now a flexibility in Congress unknown in the past, but the likelihood is that it is only a flexibility on the side of cutting taxes and increasing spending, the politically popular side. It would be surprising indeed to find Congress taking the reverse action with the same speed, no matter how clear the evidence that prompt action of that kind was essential to the stability of the economy. The deliberations might once again run into years. If this observation is correct, the conclusion that follows is that, at least for the cases in which spending cuts and/or tax increases are called for, fiscal policy will remain far less effective as a means of achieving short-run stabilization than it might be unless greater fiscal flexibility is provided. This could be done through either the formula approach with limited discretion or through an outright grant of some discretion to the President.

Variations in Government Purchases, Transfer Payments, and Taxes

We noted earlier that to a limited extent built-in flexibility provides appropriately timed, helpful fiscal response to recessionary and inflationary developments. In order to secure greater fiscal response, flexibility by formula or discretionary action is required. This brings us to the question of whether federal expenditures and tax receipts, even with discretionary action, can in practice be varied by the large amounts that may at times be necessary for successful stabilization policy.

The answer to this question is certainly quite different today from what it would have been had it been asked in those days before fiscal policy had become a generally accepted means of achieving stabilization. For example, federal government purchases made up somewhat over 1 percent of GNP in 1929—$1.4 billion out of GNP of $103.4 billion. This percentage rose during the thirties, reached almost 40 percent during the years of World War II, and for the decade 1967–76 averaged 9.0 percent. Thus, to the extent that purchases of the federal government can be promptly reduced by, say, 5 percent as an anti-inflationary measure or raised 5 percent as an anti-recessionary measure, the impact on the economy will be far greater today than a similar percentage change would have been in the days of relatively smaller federal budgets. As federal purchases have grown in importance relative to GNP, so have federal transfer payments. These rose from less than $1 billion in 1929 to almost $185 billion in 1976, although GNP in the meantime had increased less than seventeenfold. Roughly paralleling the growth of both types of federal spending, there has been, of course, a growth in federal tax receipts.

Although the monetarists stand in opposition, most economists believe that variations in government purchases, transfer payments, and tax receipts can be used in various combinations to produce desired expansionary or contractionary effects on aggregate demand and the level of income. In what follows we accept this majority view and go on to take a brief look at some of the practical limitations on the use of variations in spending and taxing as tools of fiscal policy.

Purchases of Goods and Services One way of reducing excess-demand inflationary pressures is to reduce the level of government purchases, thereby releasing resources to meet private demands. Because it looms so large in the total, the most likely candidate for reduction would appear to be purchases for national defense. Although the fraction of total federal government purchases made up of purchases for national defense ran much higher during the peak years of the Vietnam war, for the years 1974–76 it has run a little over two-thirds. Much was heard from candidate Carter in 1976 about cutting $5 to $7 billions of waste from defense expenditures, but what was heard from President Carter in early 1977 seemed to say that that would come in 1978 or later. However, whatever the level at which the defense budget is set in any year, what is relevant here is that it is unlikely that it would be deliberately cut below this level for stabilization purposes. In other words, people will differ as to what level of purchases is required for adequate defense, but once that level is decided on, few will argue that we should cut below it as a means of meeting inflationary pressures. Although almost all of the defense budget is subject to annual review through the normal appropriations process, the figure that emerges each year is thus one that reflects predominantly what those in power regard as the amount needed for national defense and only to a minor degree the figure that would contribute to the needs of economic stabilization. To the extent that this is true, it would appear that variations in federal government purchases deliberately engineered for stabilization purposes would have to be focused on the one-third of purchases that are of a nondefense nature.

However, there are problems here also. For example, sharp slashes in these expenditures are plainly not administratively feasible in the short run. Beyond this is the injustice of placing the major burden of fiscal adjustment on the nonmilitary public sector. While it may be possible to reduce or stretch out some kinds of nondefense purchases of goods and services, those that are aimed at meeting such urgent problems as urban blight and pollution control should not be subject to cutting or stretching out. It also should be noted that many of these expenditures are for services of government employees, so that adjusting spending of this kind to meet inflation means discharging workers who are performing worthwhile services. In times of inflationary pressures, it therefore does not seem that any great effect can be realized via cuts in nondefense purchases except at a high social cost. Inflation has to be attacked primarily through increases in tax rates, which places the major burden of the stabilization policy on the almost 80 percent of GNP that represents private uses of output rather than on the 2.6 percent of GNP that represents federal nondefense uses of output in the last few years.

One way of combating a recession is to increase the level of government purchases, thereby increasing government demand for goods and services and absorbing idle resources in their production. Unless there are to be more government purchases just for the sake of purchases (i.e., disregarding the usefulness of what is purchased), the major part of expanded purchases to meet the problem of recession will have to be for public works such as roads, dams, public buildings, and the like. This brings us to the problem of the limited flexibility of public works projects.[20] Even with preplanning, some lag is unavoidable between the decision to undertake a project and the actual initiation of expenditures on it. Furthermore, few public works projects are of a type that can be completed in a matter of months or even a year. Consequently, there is the possibility that the economy will recover and even enter a vigorous expansion just when many antirecessionary public works projects are half completed. To abandon them at this point would be wasteful; to complete them would accelerate the expansionary movement at a time when this would be undesirable. This lack of flexibility is not a serious problem in the face of a prolonged depression such as the one during the thirties, but it does mean that public works projects are of limited value in coping with short, cyclical downturns of the type the economy has suffered in the postwar period.

Transfer Payments Transfer payments by the federal government would appear to allow greater maneuverability than purchases, because they can be more quickly expanded or contracted as conditions require. But we encounter problems here too. To achieve a substantial reduction in transfer payments as a part of an anti-inflationary program would mean cutting benefit provisions under old age, survivors, and disability insurance, unemployment insurance, and retirement programs (which account for almost all federal government transfer payments apart from Medicare and Medicaid). Some of these payments are contractual obligations of government and cannot be touched; others, such as old age benefits, although subject to change by congressional action, occupy a place in the social fabric of this country that effectively rules out reductions. Like certain kinds of federal government purchases, certain kinds of transfer payments are in practice rela-

[20]See, for example, S. Maisel, "Varying Public Construction and Housing to Promote Economic Stability," in Joint Economic Committee, *Federal Expenditure Policy for Economic Growth and Stability*, Papers, 1957, pp. 382–97. See also his "Timing and Flexibility of a Public Works Program," in *Review of Economics and Statistics*, May 1949, pp. 147–52; J. Margolis, "Public Works and Economic Stability," in *Journal of Political Economy*, Aug. 1949, pp. 293–303; and R. L. Teigen, "The Effectiveness of Public Works as a Stabilization Device," in W. L. Smith and R. L. Teigen, eds., *Readings in Money, National Income and Stabilization Policy*, 3rd ed., Irwin, 1974, pp. 305–10.

tively uncontrollable outlays.[21] Because so much of the total of transfer payments falls into this category, sizable cutbacks in transfer payments as a means of reducing aggregate demand appear to be unattainable in practice.[22]

On the other hand, expansion of transfer payments as a means of stimulating the economy appears to be subject to no such limitation. What is more, because a large part of such funds is received by low-income persons, the likelihood is that most of such funds will be used promptly to finance an increase in consumption spending. Although transfer payments have this significant advantage on the side of stimulation, any increases in transfer payments are likely to be permanent, presenting a problem if restrictive action is subsequently indicated. This is the problem referred to in the preceding paragraph, a problem that limits the usefulness of transfer payments for short-run stabilization purposes that at times require expansionary action to be followed before long by action in the opposite direction.

Tax Receipts Fiscal policy may be used to attack the problem of excessive or deficient aggregate demand from the tax side as well as from the side of government purchases and transfer payments. For example, an overheated economy may be cooled by an appropriate rise in tax rates with no change in government spending. The only limitation to the use of tax rates for this purpose is the willingness of Congress to impose higher rates, a willingness often absent except in cases of extreme emergency, such as wartime. But, even if this reluctance to raise taxes is overcome, there is the complex problem of deciding which rates are to be raised and the amount by which each is to be raised. An overly restrictive tax policy may not only bring an inflationary expansion to an end but precipitate a decline. In short, it may not only cool an overheated economy but "freeze" it.

To the extent that excessive aggregate demand can be attributed to developments in a particular sector of the economy, it may be possible to direct tax policy toward this sector without putting the brakes on the system as a whole. For example, if a boom in investment spending is under way, a rise in corporate income tax rates with unchanged personal income tax rates may be in order. In such a case, the dampening effect will fall, at least initially, on the sector that needs dampening.[23] If, on the other hand, the excess is primarily the result of a surge in consumption spending, the personal income tax would probably be a better vehicle through which to effect the required degree of restraint.

When the economy faces deficient aggregate demand and recession, appropriate fiscal pol-

[21]For an analysis of the meaning and measurement of controllability, see Chapter 7, Budget Controllability and Planning, in B. M. Blechman, E. M. Gramlich, and R. W. Hartman, *Setting National Priorities—The 1976 Budget*, The Brookings Institution, 1975. What is controllable may be defined more or less narrowly, and on the basis of their definition the authors found only a very small percentage of the 1976 planned budget outlays of $349.4 billion to be controllable. Among domestic outlays, there was a mere $2 to $3 billion, barring a policy decision to permanently cut back on grants-in-aid and research. Among outlays for defense and international affairs, the figure was a mere $3 to $4 billion, with another $10 billion included if substantial policy shifts were made to halt new weapons programs and foreign aid commitments.

[22]We refer, of course, to discretionary reductions. Aggregate unemployment benefits will be reduced automatically as unemployment falls during the expansion phase of the business cycle. As we saw, this results from the operation of the unemployment insurance program as a built-in stabilizer.

[23]The decrease in after-tax profits of corporations may not, however, restrict the funds available to corporations to finance an expanding rate of investment spending. For example, if the outlook for growing corporate profits remains sufficiently favorable, corporations, despite the fact that government is now taking a larger share of this total, may offset the restraint of higher taxes by reducing the share of after-tax profits paid out in dividends. In addition, they may resort to borrowing as another source of additional investment funds. However, when the boom is being fed by investment spending, a restrictive monetary policy can play an important role by reducing the availability and raising the cost of funds borrowed to finance the splurge of investment spending. Recall that some of these questions were discussed in Chapter 12 in connection with the role of finance as an influence on investment spending.

icy may, of course, be to cut tax rates. Here again, the question arises of how any given cut should be allocated over various types of taxes in order to get the maximum stimulative effect. In simplified form, the question is often approached as a choice between, or a combination of, tax cuts designed directly to stimulate consumption spending or investment spending. Because the cyclical fluctuations in investment spending are relatively greater than those in consumption spending, investment spending will usually be the more depressed of the two in times of recession. For this reason, some people argue that tax cuts should be aimed at encouraging investment, since the principal need is to raise the rate of investment spending in order to move the industries that are engaged in producing capital goods closer to their prosperity levels of output. Increased activity will mean rising income for consumers, from which, by way of the multiplier, will come the rise in consumption necessary to keep the upward movement rolling. Viewed from this perspective, tax cuts aimed directly at raising consumption spending will not, except after an unacceptably long lag, raise activity in the capital goods industries. This means that the immediate stimulus is not being applied where it is most needed. Others argue, however, that there is no better stimulus to investment spending than that provided by increased consumption spending. These economists claim that unless and until business persons see an increase in the rate at which goods are moving into the hands of consumers, they will be little influenced in their investment decisions by tax inducements. Consequently, for these people the maximum stimulative effect of a given amount of tax reduction will be secured through tax changes that leave more after-tax income in the hands of consumers than in the hands of corporations. The specific arguments on both sides could be examined in detail and in more precise form, but our purpose is simply to point out that this is one of the basic questions to be answered in selecting the most effective expansionary tax policy. It should also be noted that the answer need not be the same in every situation; investment-stimulating tax changes may look more promising in one recession, consumption-stimulating tax changes in another.

Whether an expansionary tax policy is to be aimed primarily at consumption or investment spending, there are a number of techniques that may be employed for either purpose. To stimulate consumption spending, the principal reliance will be on cuts in the personal income tax, but cuts in excise tax rates may also help. To the extent that prices of taxed goods fall with lower excises, unchanged money expenditures will mean an increase in the total amount of goods that can be purchased, and this will stimulate an increase in the production of goods and an expansion of employment. Within the personal income tax structure, a stimulative effect may be gained either through cuts in rates or by such changes as increases in the size of personal exemptions or larger standard deductions. Within the rate structure, the cut may be limited to the first bracket (the "basic rate"), or it may be an across-the-board cut, or it may be still another variant. If the sole objective is to obtain the maximum stimulative effect on consumption for a given reduction in tax revenue, the most effective technique will probably be a reduction in the first-bracket rate of the personal income tax. This is the only rate paid by taxpayers with the smallest taxable incomes, and these taxpayers are those most likely to devote any increase in take-home pay to additional consumption spending. Investment spending may also be stimulated through a number of tax-related techniques. The most familiar are reductions in corporate income tax rates, liberalization of depreciation regulations, and tax credits on purchases of capital goods.

For over three decades, most economists have accepted the proposition that changes in income tax rates can and should be employed to stimulate or depress consumption

and investment spending in the interest of economic stabilization. While this acceptance has by no means turned into rejection, the confidence economists once had in the efficacy of this fiscal stabilization tool has declined since the end of the sixties. That confidence perhaps reached a peak with the success achieved by the tax cut of 1964 in stimulating the economy. It declined with the apparent failure of the 10-percent surtax of 1968 to exert the contractionary effect on consumption spending that was expected from it.[24] It is interesting to note that this development understandably gave a great boost to the monetarists' stock. Although there are various other possible explanations, this development was widely accepted as a confirmation of the monetarist argument.

Another explanation advanced by some Keynesian economists is based on an application of the permanent income hypothesis. The unsuccessful 1968 tax increase was a temporary increase and was widely advertised as such, while the successful 1964 tax decrease was a permanent decrease. As we would expect from the permanent income hypothesis, a given reduction in after-tax income that results from an increase in tax rates as in 1968 but that the consumer has been told and believes will be in effect for only a year or two calls forth a very much smaller downward adjustment in his consumption over that year or two than would the same reduction in income if it were expected to be in effect indefinitely. What consumers actually did in 1968–69 was to absorb much of the increase in their tax bill by reducing saving, thereby robbing the personal income tax increase of most of its intended contractionary effect on spending.[25]

If the 1968 experience is correctly explained by the permanent income hypothesis, it raises the question of whether future temporary tax increases or decreases can be effective. On the other hand, if the correct explanation lies elsewhere, the same question of effectiveness may not arise.

Although the experience with the 1968 tax increase came as a jolt to the many economists whose analysis suggests that it did not produce anything like the contractionary effect that they expected, this experience by no means warrants abandonment of tax changes as a stabilization device. Further study may lead to the development of appropriate techniques to meet the problems encountered in cases like this. If the workings of the permanent income hypothesis do indeed make temporary income tax changes ineffective, an alternative could be the use of a federal value-added tax whose rates would be temporarily raised as an anti-inflationary measure. While such a tax would not be subject to the permanent income problem noted for a temporary increase in income tax rates, it is regressive and raises the question of equity. Thus, what may be an approach that satisfies the objective of stabilization can be ruled out by the fact that it results in what most people feel is an unfair distribution of the tax burden among different income classes or different types of income.

If short-term variations in personal income tax rates are found to be ineffective for stabilization purposes, and if a tax like the value-added tax is unacceptable on the grounds that it is inequitable, one must turn to other alternatives or to modifications that are acceptable. An approach can probably be devised that provides an acceptable compromise between eco-

[24]See A. M. Okun, "Measuring the Impact of the 1964 Tax Reduction," in W. W. Heller, editor, *Perspectives on Economic Growth*, Random House, 1968; A. Ando and E. C. Brown, "Personal Income Taxes and Consumption Following the 1964 Tax Reductions," in A. Ando et al., editors, *Studies in Economic Stabilization*, The Brookings Institution, 1968; A. M. Okun, "The Personal Tax Surcharge and Consumer Demand, 1968–70," in *Brookings Papers on Economic Activity*, 1, 1972, pp. 211–20; and W. L. Springer, "Did the 1968 Surcharge Really Work?" in *American Economic Review*, September 1975, pp. 644–59.

[25]See R. Eisner, "Fiscal and Monetary Policy Reconsidered," in *American Economic Review*, Dec. 1969, pp. 897–905.

nomic stabilization and the other goals of public policy. Because of the limitations on the use of short-term variations in government spending for stabilization purposes, the ability to make use of short-term variations in taxes is especially important. However, as we have seen in our discussion of the variables through which fiscal policy is carried out—namely, government purchases, transfer payments, and taxes—there are difficult problems associated with the use of each. The better these and other problems are solved, the more effective fiscal policy will be and the greater the contribution it can make toward achieving a more stable economy.

chapter 25
Monetary Policy

Monetary policy is the exercise of the central bank's control over the money supply as an instrument for achieving the objectives of general economic policy. To the degree that monetary policy contributes to the achievement of such objectives as full employment, stable prices, and economic growth, it does so primarily by influencing the level of aggregate demand and thereby the level of money income. The first part of this chapter is concerned with a fundamental question that emerges from this. How does monetary policy work, i.e., what is the transmission process by which a change in the money supply brings about a change in the level of money income? Although there are more than two theories of the transmission process, one may still identify a basic two-way cleavage, which is designated as Keynesianism versus monetarism. The Keynesian theory sees changes in the money supply working their way through the system in a way that does not result in a close and stable linkage between changes in the money supply and changes in the income level. The monetarists do see such a close and stable linkage. To the extent that the monetarist view is correct, money is an extremely important influence on the level of income; to the extent that the Keynesian view is correct, money is that much less important.

The difference in view as to the nature of the transmission process gives rise to a difference in view as to what are the proper guides for the monetary authorities to follow in setting monetary policy. How are they to determine when it is appropriate to adopt an easier policy—that is, to increase the rate at which they are expanding the money supply—and when it is appropriate to adopt the opposite policy? Before 1970 the Federal Reserve authorities used the movement of interest rates or, more broadly, credit conditions, or the "tone and feel" of financial markets, as their almost exclusive guide to what appropriate policy should be. For example, an upward movement in interest rates that appeared to be so large and so rapid as to threaten an existing healthy economic expansion would lead the authorities to expand the money supply more rapidly than otherwise. This approach, apart from other considerations, may be seen to be consistent with the Keynesian view of the transmission process. The monetarist view, however, is that Federal Reserve policy guided by the movement of interest rates will not contribute to stability of the economy but just the opposite. To the monetarists, the money supply or some similar monetary aggregate is the only proper kind of guide to monetary policy. That is, what the Federal Reserve

authorities should do is to take whatever actions are needed to make the money supply grow at a predetermined target rate, regardless of what may happen to interest rates in the process. This question of alternative guides to monetary policy, monetarist and traditional views, is treated in the second section of the chapter.

In the final section of the chapter, we turn once again to the question of the closeness of the relationship between the money supply and the income level. This broad question has many aspects, some of which were touched on earlier. Here we look into the way in which certain institutional and structural features of the economy operate to offset the effect of changes in the money supply and thereby to loosen the link between the money supply and the income level. The result is that the monetary authorities are less able than otherwise to bring about desired changes in the income level via money supply changes. It may appear that the influence of these features would not result in such a restraint on the authorities, because the authorities can compensate for this offsetting by simply pursuing a given policy more intensively. However, we will see that there are still other features of the economy that prevent them from taking the more extreme action that would be called for.

HOW DOES MONETARY POLICY WORK? KEYNESIANISM VERSUS MONETARISM[1]

How does monetary policy work its effect on the economy's income level? In other words, what is the transmission process by which a change in the money supply causes a change in the level of income? It is one thing to outline, as in the simple Keynesian model of earlier chapters, how an increase in the money supply will cause a decline in the interest rate and why, with a given MEC schedule, a lower interest rate will lead to a rise in investment spending and therefore to a rise in income. This is satisfactory as far as it goes, but it does not come to grips with the actual transmission process. A major development in monetary theory during the fifties and early sixties was the development of an explanation of that transmission process in terms of a systematic theory of portfolio adjustments. A change in the money supply works its way through to produce a change in the level of income by setting off a complex sequence of substitutions among the financial and real assets that make up wealth-holders' portfolios.[2]

The Portfolio Adjustment Process[3]

An explanation of the process begins with the recognition that, in a broad sense, there is a rate of return on all assets in that all assets provide their owners with benefits. This is quite apparent in the case of plant and producers' durable equipment for which a specific rate of return may be readily computed and for interest-bearing financial assets like bonds for which yields are reported each day in the pages of financial newspapers. Although less apparent, it is also true for money and goods held by

[1]Although the analysis here focuses on the basic distinction between the transmission process as found in Keynesianism and monetarism, more detailed distinctions may be made. For a survey that goes into such detail, see R. W. Spencer, "Channels of Monetary Influence: A Survey," in *Review*, Federal Reserve Bank of St. Louis, November 1974, pp. 8–26.

[2]The portfolio balance theory of monetary behavior is primarily associated with the name of James Tobin. For an outline of this approach, see his "Money, Capital, and Other Stores of Value," in *American Economic Review*, May 1961, pp. 26–37.

[3]We will here give only a minimum outline of the process. For a somewhat fuller statement, see M. Friedman and A. J. Schwartz, "Money and Business Cycles," in *Review of Economics and Statistics*, Supplement, Feb. 1963, pp. 59–63.

consumers. Goods like the family automobile and home appliances provide obvious flows of services to their owners. Less obvious but still present is the flow of services provided to its owners by money in the form of the convenience and security that immediately available purchasing power offers. Although the flows of services provided by such assets cannot be expressed as a rate of return in the way that one expresses the income from a capital good or a bond, a rate of return nonetheless exists for these other assets.

Given then that all assets have rates of return or yields, suppose now that the Federal Reserve purchases Treasury bills in the open market, which leads to a rise in the price of bills and a decline in their yield. With yields on other securities unchanged at the moment, a process of arbitrage begins in which wealth-owners execute portfolio adjustments that tend to push down yields on financial assets in general. Furthermore, the Federal Reserve's initial purchase, other things being equal, has added to the reserves of the banks, and this will lead them to purchase more securities and make more loans. Their purchases of securities will add to the downward pressure on the yields provided by these assets, and their increased supply of loan funds will make for lower interest rates to borrowers.

As far as individual wealth-holders are concerned, the initial disequilibrium created by the increase in the amount of money in their portfolios is corrected as the process of trading money for other financial assets is carried to the point at which they find no advantage in further substitution of this kind. At the relative yields now in effect, their portfolios contain that amount of money and that amount and distribution of other financial assets which provide portfolio balance. This, however, is by no means a full equilibrium. It neglects the fact that the expected yields on various real assets are initially unchanged by the increase in the money supply and by the decline in yields on financial assets in general. Although wealth-holders may, in the way described above, first tend to substitute other financial assets for what have become excess holdings of money, they will sooner or later also substitute real assets for financial assets. This constitutes an increase in the demand for real assets. Some may acquire existing assets directly by purchasing such things as apartment buildings and other commercial real estate; others will acquire ownership claims to various kinds of real assets, including plant and producers' durable equipment, by purchasing shares of stock. Increased demand for the existing stock of real assets will mean higher prices for them, and this will stimulate the production of more such goods. Increased demand for stock shares will mean higher prices, and this will encourage corporations to issue more stock to finance expansion of their productive facilities. This will lead to the production of more such goods. Similarly, corporations may choose to issue more bonds for this same purpose because the rate at which they can borrow in the bond market is now lower than it was.

Although this appears to be the more important component, the increased demand for real assets is not limited to the purchase of capital goods used in business operations or to ownership claims to such goods. The portfolio adjustment process will also lead wealth-holders to increase the demand for capital goods of a consumer type. An increase in the supply of money will bring about a series of substitutions in portfolios which when completed will mean an increase in holdings of such assets as consumer durable goods as well as of other real assets and financial assets.

Subject to an important qualification for the wealth effect, which we will consider below, it appears that the step in these portfolio adjustments at which there is an increase in the demand for real assets or a substitution out of financial assets and into goods will take place *only* as a result of changes in relative yields on different assets. *The substitution into real assets must be the result of the fact that the yields on financial assets have fallen relative to the expected yields on real assets.* If wealth-

holders are responsive to the changes in relative yields, they will, in the present example, adjust their balance sheets to include more real assets and less financial assets, an adjustment that will stimulate production of more capital goods and raise the level of income.

This proposition is fundamental because, in the judgment of at least some economists, it leads to the heart of the difference between the Keynesians and the monetarists. Although Friedman and other monetarists trace the effect of changes in the money supply through a portfolio adjustment process much like that described, they do not hold that changes in interest rates are a prerequisite to changes in the demand for goods and services. Following an increase in the money supply, there can be a portfolio adjustment involving a movement out of money directly into goods. This is suggested by the following kind of statement: "The end result need not be a change in interest rates at all; it may be a change in the general price level or in output. An increased amount of water may flow through a lake without raising its level more than momentarily."[4] An increase in the money supply can, in other words, lead directly to spending for real assets. The monetarists thus do not accept the Keynesian view that spending can be affected only indirectly as changed rates or yields on financial assets alter the prospective profitability of acquiring real assets and thereby affect the rate of spending for the various kinds of real assets.

It is one thing for monetarists to reject the Keynesian explanation; it is another to present an acceptable alternative explanation. It seems that most economists who have been able to follow the debate to the depths it reached do not believe that the monetarists have really provided a convincing alternative explanation to support the contention that a change in the money supply in and of itself can lead promptly and directly to a change in the demand for goods. In the absence of wealth effects, there appears to be no way to explain how a change in money can have a direct effect on the income level.

Consider again the case in which the Federal Reserve purchases U.S. government securities in the open market. As it bids up the prices of government securities in order to persuade wealth-holders to exchange some of their holdings for new deposits, it effects a change in the composition of their assets but not at this point in their wealth or income. In bidding up prices, it also brings about a fall in the yields on government securities. If the wealth-owners in question now choose to convert some of their deposits into goods, it must be because the amount of goods they wish to hold, like the amount of money they wish to hold, is affected by the change in the yields on government securities and similar financial assets. Recall that the lower yield on government securities will, via a process of arbitrage, lead to lower yields on other securities. At the lower yield on securities, portfolio equilibrium calls for substitution involving not only more money and less securities but also more goods and less securities. Without the change in yields, there seems to be no explanation of why the demand for goods increases, but with it we do have such an explanation. However, it is an explanation that shows that the connection between changes in the supply of money and the level of income is an indirect one. And if this is the only explanation, it follows that there is no direct connection as suggested by the monetarists.

Set forth in this manner, this conclusion seems to fly in the face of common sense. After all, to noneconomists the monetarist conclusion of a direct relation is one of the most self-evident propositions in all of economics. As they see it, since an increase in the amount of money means that the public as a whole now has more money than before, such an increase must mean the public will raise its total spending above what it was before. An increase in

[4] M. Friedman and D. Meiselman, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in *Stabilization Policies*, Commission on Money and Credit, Prentice-Hall, 1963, p. 221.

the money supply from one period to the next is, after all, that much additional "disposable income." There is an unquestioned direct relationship between a change in disposable income and the level of demand, so there must be the same between a change in the money supply and the level of demand.

Changes in Money and Changes in Wealth

Not even the staunchest antimonetarist denies that there may indeed be a relation of the kind just described, but he will hold that this relation is not a unique result of the fact that there has been a change in the money supply but is rather a result of the fact that the change in the money supply involved a change in the public's total wealth. He will hold that, in the absence of a change in wealth, a change in the money supply has no direct effect on the demand for goods. Accordingly, whether or not a change in the money supply brings about a change in wealth becomes a very important question. The answer is that some changes in the money supply do and others do not.

Because an increase in the money supply will tend to cause a decline in interest rates, it thereby tends to cause an increase in the discounted value of the streams of income yielded by assets. Assets whose values are determined by capitalizing the income streams they produce will rise in value when those streams are capitalized at a lower interest rate. Then as the public sees the market value of its assets grow larger by way of these capital gains, it may be expected to increase its demand for goods and services. This is the interest-induced wealth effect discussed in Chapter 9 in connection with consumption demand. It is one way in which a link may be established between changes in the stock of money and the level of demand that does not depend on *relative* interest rates.

While it cannot be denied that this interest-induced wealth effect is present, its strength, like that of the price-induced wealth effect or Pigou effect, is open to question. There is another and perhaps potentially much more important wealth effect that *sometimes* accompanies a change in the money supply. This is an effect that the general public believes always accompanies an increase in the money supply, because it believes that the creation of more money, in and of itself, adds that much to the public's wealth. Unlike other things, it is assumed that an increase in the total money supply is something the public gets without giving up anything in exchange. The fact is that this is simply not true of most of the increases that occur in the money supply—they do not bring about any such direct increase in the public's wealth.

Take first the case of an increase in the money supply that comes about as the Federal Reserve purchases securities in the open market. Assuming that the sellers are not banks, the result is that the public now holds more money and less securities. There is a change in the composition of the public's assets, but no change in its total assets or total liabilities and thus no change in its wealth. The wealth-holders in question shifted out of securities and into money because of the attractive price at which they could sell the securities, but the mere fact that they now hold more money is no reason to expect them to proceed to switch out of money into goods. However, as we saw above, such a switch may indeed follow, not because of the increase in money holdings but because of the changes in relative yields that result therefrom, the unchanged yields on real assets or goods now being relatively attractive in comparison with the lower yields on financial assets.

Unlike the wealth-holders in the first case who secured additional money by giving up securities for it, consider now wealth-holders who secure additional money by borrowing at the commercial banks. Other things being equal, Federal Reserve Bank or commercial bank purchases of securities in the open market and commercial bank purchases of customer's promissory notes result in an increase in the economy's total money supply. The cases

differ, however, in this respect. Wealth-holders who were induced to switch out of securities into money are under no special pressure to use the money immediately. Persons and firms who secure additional money by borrowing at the banks are, however, quite certain to make immediate use of that money. Why else, one may ask, would they go into debt to get it? The interest rate they must pay on the debt incurred is almost sure to be higher than the rate of return on holdings of demand deposits. Some borrowers may have incurred debt to pay off other debt, but many will do so to finance the purchase of goods. Thus, there is a presumption that an increase in the supply of money that results from an expansion of loans by the commercial banks will be promptly put to use and that it will lead quite directly to an increase in the demand for goods.[5]

Although money created by loans is thus likely to have a greater effect on the demand for goods than money created by the monetization of existing financial assets such as U.S. Government securities, the effect in both cases is achieved essentially via changes in relative interest rates or yields. In neither case can one argue that there is a direct effect on the demand for goods that arises from a wealth or income effect. In these cases, increasing or decreasing the money supply does not mean adding to or subtracting from the public's wealth or income.

This brings us to the case in which an increase or decrease in the money supply may be expected to have a direct effect on the demand for goods by adding that much to or subtracting that much from the wealth of the public. We will have such a case in the event that the public increases or decreases its money holdings without giving up or gaining an equal amount of other assets or without incurring an equal increase or decrease in its liabilities. This result will come about only in the event that the U.S. Treasury finances a U.S. Government deficit in a way that increases the money supply. There will be such an increase in the money supply if the Treasury meets the deficit by issuing newly printed currency or by obtaining newly created deposits at the Federal Reserve Banks or the commercial banks in exchange for Treasury interest-bearing obligations. As the Treasury pays out the currency to the public or transfers these deposits to the public, the public will find itself with an increase in its money holdings which, from its point of view, is a net increase in its financial assets or in its wealth. This in turn gives rise to a wealth-induced increase in the demand for goods.

Keynesians and monetarists agree that a change in the money supply that comes about in this way will produce a direct change in spending on goods via the change in the public's wealth that it produces. However, apart from whatever interest-induced wealth effect may result, this is not true for all those changes in the money supply that occur as the banking system makes loans to or purchases securities from the public. Here the Keynesians insist that the effect is an indirect one that works through portfolio adjustments in the way described above, but they also grant that one effect of an increase in the money supply will in this way be an increase in the income level.[6]

[5]Although it is true that, quite apart from interest rates charged, increased availability of loans at banks will in itself lead to loans that otherwise would not have been made, interest rates charged relative to other rates or yields still play the same kind of role here that was described earlier. Portfolio adjustments in response to changes in relative yields include more than substitutions among an existing total of assets; they also include varying the amount of liabilities and thereby the total amount of assets. In the present case, a decline in interest rates charged by banks on loans, with no decline in the prospective rates of return on the real assets to be acquired with the borrowed funds, explains why many loans are made. From an initial portfolio equilibrium, the public will be induced to enlarge its assets and liabilities by borrowing if the rates paid on borrowed funds relative to the rates expected on the goods purchased with these funds are such as to make these changes in portfolios advantageous.

[6]This was not always true. As late as the fifties, ultra-Keynesians seemed to believe they were living in a liquidity-trap world in which ordinary monetary policy was powerless to affect interest rates and income levels. No one nowadays denies that changes in the money supply will affect these variables.

Looked at in this way, it almost seems that the difference between the Keynesians and monetarists is not in whether changes in the money supply affect the income level but in how they affect it. There is some truth in this observation, but going one step further reveals that the way in which changes in the money supply work their effect on the income level makes a critical difference in terms of how close and how stable the relationship is between changes in the money supply and the income level. The Keynesian approach finds the relation to be a loose one or one subject to considerable variation over time; the monetarists maintain that the relation is fairly close or one subject to only moderate variation over time. This brings us to the question of the source of this difference.

Substitutability among Assets

The Keynesians and the monetarists present very similar descriptions of the process of portfolio adjustment, but they disagree on a critical aspect of this process: the closeness of substitution between money and other financial assets and between money and real assets. To the Keynesians, it is money and other financial assets that are close substitutes. Accordingly, because a rise in the price of any good will increase the demand for another good that is a close substitute, the rise in the prices of other financial assets, here caused by an increase in the money supply, will produce an increase in the amount of money wealth-owners wish to hold. But at the same time that higher prices and lower yields on other financial assets will mean that the public wants to hold more money as a substitute for those assets, it also means that the public will want to hold more real assets. In the way sketched earlier, the increase in the demand for real assets will lead to an increase in the production of goods and so to an increase in the level of income. But what we have here is a loose connection between a change in the money supply and the change in the level of income. Depending on the elasticities involved, a given increase in the money supply can give rise to a wide range of possible changes in the income level.

This is not so in the monetarist theory or at least not in Friedman's version of it. In his theory it is real assets and not financial assets that are close substitutes for money. An increase in the money supply sets into motion a process of portfolio adjustment, but somehow through all the substituting that takes place in the process, there is no substitution of money for financial assets. That is, despite lower rates of return on other financial assets, the public does not choose to hold more money and less of other financial assets. The demand for money is not interest elastic. What the public does choose to do is to hold more real assets or goods and less money or to substitute real assets for money. Again the increased demand for real assets that arises in this way will lead to an increase in the production of these assets and thus to an increase in the level of income. But unlike the Keynesian theory in which the public seeks to substitute goods for only a part of any change in the money supply, and a part that can vary from time to time, in Friedmanian theory substitution of goods for money will continue until the rate of production of goods or the level of income is such that the amount of money actually held is equal to the constant proportion of income that the public wants its money holdings to equal. Otherwise expressed, at this point all of the money actually held by the public will be required to mediate the volume of transactions associated with that higher level of income.

The crux of the difference between Keynesian and Friedmanian theory seems to be found in this difference between Keynesians and Friedman on the matter of substitutability among assets. From this difference flows Friedman's quantity theory conclusions and the Keynesians' rejection of those conclusions. In terms of the elementary formulation employed in Chap-

ter 15, if one denies that there is a substitution between money and other financial assets as their relative yields change, he finds that there will be equilibrium between the nominal supply of money, M_s, and the demand for money when $M_s = P \cdot k(Y)$. If one accepts that there is substitution between money and other financial assets when their relative yields change, he finds that there will be equilibrium when $M_s = P \cdot k(Y) + P \cdot h(r)$. The first equation yields the quantity theory conclusion of a stable relationship between M_s and $P \cdot Y$; the latter yields no such relationship. Since such a relationship is the essence of monetarism, the issue of the closeness of the substitution between money and financial versus real assets is crucial. If real assets are closer substitutes for money then other financial assets, we are led toward the quantity theory; if the other way around, we are led toward the Keynesian theory or toward the rejection of the quantity theory.

Until the seventies, there was a fairly general agreement among economists that the crux of the difference between the sides was the one here described. Friedman had not provided a theoretical framework of a kind that would have clearly brought out the difference that existed between himself and the Keynesians, but his writings suggested that the difference over the interest elasticity of the demand for money was central. In answer to the critics who kept asking for a theoretical framework that would explain the way in which money produced the effects on income that Friedman alleges the empirical record shows, Friedman published a major article in 1970 designed to do this.[7] Perhaps in the attempt to improve communication with other economists who are accustomed to thinking in these terms, he made use of the *IS–LM* apparatus in this framework.

The early reaction to Friedman's framework was one of confusion and consternation among leading monetary economists on the monetarist side as well as the Keynesian side. For example, James Tobin stated:

> I have been very surprised to learn what Professor Friedman regards as his crucial theoretical differences from the neo-Keynesians. . . . First, let me explain what I thought the main issue was. In terms of the Hicksian *IS–LM* language . . . I thought (and I still think) it was the shape of the *LM* locus.[8]

However, in the 1970 article, Friedman explicitly disavows his earlier belief that the demand for money is interest inelastic and thereby accepts the Keynesian-shaped *LM* curve that slopes upward to the right. Apparently no less an economist than James Tobin had come to believe that the foundation for Friedman's monetarist propositions was a vertical or near ver-

[7]For years Friedman chose to support his monetarist propositions solely on the empirical regularities he detected in his extensive studies of the historical relation between changes in the money supply and changes in the income level. In connection with the assertion that monetary changes are the key to major movements in money income, he and Anna J. Schwartz wrote, "We have great confidence in this assertion. We have little confidence in our knowledge of the transmission mechanism, except in such broad and vague terms as to constitute little more than an impressionistic representation rather than an engineering blueprint." ("Money and Business Cycles," *op. cit.*, p. 54.) This great confidence that money is the key rested on evidence of various kinds: historical case studies, e.g., the monumental *Monetary History of the United States, 1867–1960*, National Bureau of Economic Research, Princeton Univ. Press, 1963, by Friedman and Anna J. Schwartz; summary regressions of time series in which changes in income are regressed against changes in money, e.g., M. Friedman and D. Meiselman, *op. cit.*; and studies of the timing relationship between changes of the money supply (or changes in the rate of change in the money supply) and changes in the level of money income, e.g., Friedman's "The Lag in the Effect of Monetary Policy," in *Journal of Political Economy*, Oct. 1961, pp. 447–66. This impressive body of empirical evidence has been subjected to much criticism. For example, on the matter of timing relationships, which are so crucial, see J. Kareken and R. Solow, "Lags in Monetary Policy," in *Stabilization Policies*, Commission on Money and Credit, Prentice-Hall, 1963, pp. 14–25, and J. Tobin, "Money and Income: Post Hoc Ergo Propter Hoc?" in *Quarterly Journal of Economics*, May 1970, pp. 301–17. There was clearly a real need for Friedman to provide a theoretical framework to supplement his voluminous empirical work, and this was done in "A Theoretical Framework for Monetary Analysis," in *Journal of Political Economy*, March–April 1970, pp. 193–238, and a follow-up article, "A Monetary Theory of Nominal Income," *ibid.*, March–April 1971, pp. 323–37.

[8]J. Tobin, "Friedman's Theoretical Framework," in *Journal of Political Economy*, Sept.–Oct. 1972, p. 853.

tical *LM* curve. Now we are told that these propositions do not rest at all on this ground, and many economists, including the experts, have great trouble trying to detect what ground, if any, they do rest on.

Two leading monetarists, K. Brunner and A. H. Meltzer, whose work is otherwise fairly close to Friedman's, find that Friedman's framework is in important ways neo-Keynesian. In their words, Friedman's "view of the transmission mechanism [pp. 216–17] brings him into general agreement with the neo-Keynesians about the transmission of monetary policy. . . . We regard Friedman's discussion as either misleading or a complete reversal of his often stated position."[9] For these and some of his other critics, Friedman has replies.[10]

While the debate has by no means ended, the last five years have not seen another major exchange between the two sides like that of the early seventies noted here. The debate between Keynesianism and monetarism is a wide-ranging one, and what has been covered in this section is one aspect of that debate. However, this aspect—the difference in view as to the way that monetary policy works—is a basic one because it gets into the question of the importance of changes in the money supply as a determinant of changes in the income level. The monetarists persist in their belief that money is the key determinant of such changes, and today's Keynesians are equally insistent that it plays nothing like the decisive role ascribed to it by the monetarists.

GUIDES TO MONETARY POLICY

The differences between monetarists and other economists as to the way that monetary policy works is related to the differences that exist between them as to the guide or guides to follow in determining monetary policy. Knowing that monetary policy is carried out by varying the money supply obviously tells us that a change in policy calls for a change in the money supply, but this tells us nothing about what monetary policy should specifically be at any time. What is needed is a guide or guides to direct those whose responsibility it is to decide this question.

We earlier noted the major goals of macroeconomic policy—full employment, price stability, and satisfactory growth—and one might argue that these should be the guides to follow in setting monetary policy. For example, as the economy suffers a downturn and the unemployment percentage rises, monetary policy should become expansionary; i.e., the authorities should carry out those open-market purchases and/or take other steps that will increase the rate at which the economy's money supply is growing. While it may seem at first glance that this is obviously the appropriate policy, taking a so-called ultimate goal variable like full employment as a guide to policy is impractical for a number of reasons. In trying to determine how rapidly to expand the money supply in the case of a downturn (or how rapidly to contract it in the case of inflation), the monetary authorities cannot base their decision on the behavior of a variable like the unemployment percentage, which is importantly affected by various forces other than monetary policy. In othe words, to take the unemployment percentage as the guide and pump into the system whatever amount of additional money is needed to correct the unemployment problem could be disastrous; it could produce serious inflation without even then fully correcting the unemployment problem.

[9] "Friedman's Monetary Theory," in *Journal of Political Economy*, Sept.–Oct., 1972, p. 846.

[10] The September–October 1972 issue of the *Journal of Political Economy* also includes critical articles by D. Patinkin, pp. 883–905, and P. Davidson, pp. 864–82, and "Comments on the Critics" by Friedman, pp. 906–50.

Furthermore, the ultimate goal variables are affected not only by forces other than monetary policy but also by monetary policy only with some lag, and, according to some economists like Friedman, a lag that is, on the average, not only long but quite variable from one case to the next. For this reason, even if variables like output, employment, and prices were completely determined by monetary policy, which is not at all the case, they would still not be a satisfactory guide. They would always be indicating the results of monetary policy pursued some time earlier. What is needed as a guide is a variable that promptly responds to monetary policy so that the authorities can tell from it the direction in which their most recent policy decisions have been carrying them.

One more difficulty of this kind is that changes in the overall position of the economy that are reflected by changes in the unemployment percentage, price indexes, and growth rates are observable only after some time. If the monetary authorities await clear signs that a downturn is under way before they adjust policy to meet the problem, the problem will be more difficult to handle than if it had been attacked sooner. However, if they move to an expansionary policy at the first sign of a downturn, they may create a problem where none existed before. For if the first sign was a false sign, then expansionary policy will not check a downturn since that is not occurring, but it will produce undesirable inflationary pressures that would not have existed otherwise.

Yet another problem, of a different nature, that is faced in trying to use the ultimate goals as guides is the incompatibility that exists among some of them, e.g., between full employment and price stability. Which way should monetary policy turn if it bases policy on these ultimate goals but finds itself faced with a rising unemployment percentage and a rising rate of inflation, an extreme case of which was faced in 1974?

Because of these various difficulties, the monetary authorities must choose as a guide a variable that is much more closely tied to monetary policy in the sense that changes in its value are substantially determined by changes in monetary policy and promptly follow them. Such a variable provides the authorities with an indicator of the direction in which and speed with which they are moving. It is also, of course, indispensable that this variable be related to the ultimate goal variables if it is to be of any value as a guide. That is, changes in it must be expected to lead to changes in these goal variables. The authorities can then operate to produce changes in the chosen variable with some assurance that these changes will work through the system to produce changes in those ultimate variables whose control is the end goal of monetary policy.[11]

Monetary Aggregates versus Credit Conditions as Guides

What variables meet these requirements? As a rule, we find that a variable that comes closer to meeting one requirement falls farther short of meeting the other one. Consider the following illustrations. A variable that changes only because of actions taken by the Federal Reserve authorities is the amount of U.S. government securities held by the Federal Reserve Banks. However, this variable is not at all closely related to the ultimate goal variables, because short-run changes in the Reserve Banks' security holdings may occur without causing changes in member bank reserves or the public's currency holdings and thereby without causing changes in the money supply. And it is through changes in the money supply, either directly or via interest rates, that the monetary authorities are able to influence the ultimate goal variables.

From this example one might conclude that

[11] For a full discussion of the various aspects of this complicated area of guides, targets, indicators, and goals of monetary policy, see the various papers in K. Brunner, ed., *Targets and Indicators of Monetary Policy*, Chandler, 1969, and T. R. Saving, "Monetary-Policy Targets and Indicators," in *Journal of Political Economy*, Aug. 1967, Part II, pp. 446–65. See also B. M. Friedman, "Targets, Instruments, and Indicators of Monetary Policy," in *Journal of Monetary Economics*, October 1975, pp. 443–73.

the monetary authorities should choose the money supply as their guide variable. While it is argued, at least by the monetarists, that the money supply shows a relatively close and stable relationship with the ultimate goal variables, it is not a variable whose changes can promptly and accurately indicate to the monetary authorities the impact of the actual actions taken by them. This is because short-run changes in the money supply are not completely determined by Federal Reserve actions. Various other factors such as changes in the public's currency-to-deposit ratio or in the commerical banks' excess reserves-to-deposit ratio affect this. Thus, a rising money supply does not necessarily tell the Federal Reserve authorities that their actions on balance have been expansionary—it is possible that the money supply is rising in spite of actions taken by the Federal Reserve authorities, not because of them. If the authorities evaluate the direction and speed with which they are moving by the changes that occur in the money supply, the evaluation of their policy stance may at times involve sizable error.

A possible alternative guide variable is the amount of reserves held by the commerical banking system. Changes in this variable are almost completely under Federal Reserve control, so they reflect changes in Federal Reserve actions more closely than changes in the money supply do, but because changes in the amount of reserves do not mean proportional changes in the money supply, they are less closely related to the ultimate goal variables.

The nature of the problem may be apparent from these illustrations. Without going into it any further, we may see that a number of related variables like commercial bank reserves, the monetary base, the money supply, or total commercial bank credit are all possible guides to monetary policy but none ideally meets both of the requirements noted. In the present context this class of variables is usually referred to as *monetary aggregates*.

A quite different class of variables that can serve as guides are those that reflect conditions in credit markets. Credit conditions are measured by variables like the amount of member bank borrowing from the Federal Reserve Banks, the free reserve position of the member banks, the Treasury bill rate, the Federal funds rate, and other interest rates, and the intangible known as the "tone and feel" of the market. Among these variables, interest rates receive the most attention, and we will focus on them in what follows.

The question arises as to whether or not interest rates meet the requirements set forth above for a guide to monetary policy. Whatever the extent of their control over a monetary aggregate like the money supply, the Federal Reserve authorities have less control over interest rates. However, they are able to bring about short-run marginal changes in interest rates, and this is what is significant here. But because interest rates do not change solely in response to actions taken by the Federal Reserve authorities, it is not possible for the authorities to accurately evaluate what their policy has been by observing the movement of interest rates. Thus, falling interest rates do not necessarily tell the authorities that their actions on balance have been expansionary—it is possible that interest rates are falling not because of actions taken by the authorities but in spite of them. The room for error here is greater than in the use of the money supply as a guide, because changes in the money supply are tied to changes in Federal Reserve actions more closely than changes in interest rates are. With regard to the other requirement—i.e., that the variable chosen as a guide be one whose changes lead to changes in the ultimate goal variables—economists who believe that monetary policy exerts its effect on those ultimate goal variables by bringing about changes in interest rates obviously give a high rating to interest rates as a guide.

The question of what guide should be followed by the monetary authorities is essentially one of choosing between a monetary aggregate like total reserves or the total money supply and a measure of credit conditions like interest

rates. Prior to 1970 the Federal Reserve authorities had been guided in policy-making almost completely by credit conditions, especially as reflected by interest rates. Starting in 1970 they began to pay far more attention to the money supply and other aggregates than they had before, and this has continued over the following years, although the relative emphasis placed on each as a guide to policy is not always clear.

The choice that is made between the two guides, or the choice of a compromise that puts greater emphasis on one or the other, reflects in part the differences in policy-makers' beliefs as to the way the transmission process works. Earlier in this chapter we traced the Keynesian argument that, apart from wealth effects, monetary policy works only by effecting changes in interest rates or yields and the monetarist counterargument that monetary policy can work directly. What difference does this disagreement make in the choice of the more appropriate guide to monetary policy? To try to answer this question, let us take as a clear-cut illustration an economy operating at a high level of employment and with strong inflationary pressures at work. Then the appropriate monetary policy would surely appear to be a restrictive one or one that reduces the rate of growth of the money supply. While this might readily be granted, the next question is how restrictive the policy should be. To those who believe that the effect of changes in the money supply are transmitted through changes in interest rates, the guide as to how far to push a restrictive monetary policy is provided by the behavior of interest rates. Rates may not be allowed to rise too rapidly or to too high a level or a downturn may result, but they must be made to rise at some appropriate speed and to some higher level to get the restrictive effect that is desired. Although they grant that many other variables must be taken into account, those who see the transmission process in this way maintain that the primary guide to the authorities in deciding on whether policy should be made more or less restrictive as the situation develops is the movement of interest rates.

The major problem encountered in employing interest rates as a guide to policy is that of distinguishing between two kinds of changes in interest rates: those that are due to temporary or irregular changes in the demand for funds and those that result from the first stage of a sustained change in the demand for funds that springs from an underlying major change in aggregate spending. The first kind can properly be met with an appropriate change in the money supply, but the second kind should not automatically be resisted. To offset the first is consistent with the central bank's traditional concern for the stability of financial institutions and money markets and not inconsistent with the overall stability of the economy, but to offset the second may not be consistent with overall economic stability. Although it may mean fairly rapid and large changes in interest rates, with the shock this can have on financial institutions and markets, the second kind of change is the one that should be allowed to proceed without resistance because it contributes to overall economic stability.

This may be seen as follows. In an economy operating near full employment, if the monetary authorities resist an upward movement of interest rates that results from an upward surge in aggregate demand, the policy of stabilizing interest rates causes destabilizing changes in the rate of growth of the money supply. The rapid increase in the money supply that results from their effort to hold down interest rates will, after some lag, reinforce the rise in spending and generate inflation. The final result of adhering to the interest rate guide in such a case is then to add to whatever instability arises in the economy for other reasons rather than to offset it. Instead of monetary policy being countercyclical, it turns out to be procyclical. Not to have resisted by turning to an expansionary monetary policy would have permitted the rise in interest rates to do what it should do in a case like this—choke off the rise in demand and thus prevent the inflation to which it would otherwise lead.

The Federal Reserve would not have resisted

a rise in interest rates in a case like this nor a decline in a case of the opposite kind if the change were recognized as this kind of change. But the problem is that the Federal Reserve cannot be sure whether any given movement in interest rates is temporary and reversible or sustained and irreversible until the movement has persisted for some time, and by that time a great deal of damage may be done if it seeks to prevent the changes that would otherwise occur. Because of its traditional concern with the stability of financial markets, the Federal Reserve has often moved to prevent interest rates from changing when they should have been permitted to do so. An illustration is found in the 1965 escalation of the Vietnam War, which saw rapidly rising military expenditures added to an already vigorously expanding economy. Under the circumstances the Federal Reserve authorities might have been expected to adopt a policy of slowing the rate of expansion of the money supply, but they actually did the opposite. The policy they followed reflected their concern over the fact that interest rates were rising quite rapidly during the year. It was not until December that they recognized that the pressures on interest rates were not temporary. Their reaction was to reverse policy in 1966, which replaced the rapid rate of money growth with a rate approaching and, during part of the year, actually reaching zero. Interest rates in 1966 then showed one of the largest jumps on record. There was much disorder in financial markets, culminating in a "credit crunch" by late summer of 1966. A few months later the Federal Reserve authorities once again reversed policy.

Critics of the Federal Reserve argue that the emphasis placed on stability of interest rates by the authorities results in errors in both the timing and the scope of action. Using the 1965–66 episode as an example, Friedman wrote as follows in 1968:

> Too late and too much has been the general practice. For example, in early 1966, it was the right policy for the Federal Reserve to move in a less expansionary direction—though it should have done so at least a year earlier. But when it moved, it went too far, producing the sharpest change in the rate of monetary growth of the post-war era. Again, having gone too far, it was the right policy for the Fed to reverse course at the end of 1966. But again it went too far, not only restoring but exceeding the earlier excessive rate of monetary growth. And this episode is no exception. Time and again this has been the course followed—as in 1919 and 1920, in 1937 and 1938, in 1953 and 1954, in 1959 and 1960.[12]

This statement, if written ten years later, would have added to this enumeration the episodes of 1969–70 and 1974–75. The objective of policy in 1969 was to slow the booming economy, the hope being to reduce the rate of inflation without causing a recession. The growth of the money supply was cut back in the last half of 1969 to a 1.2-percent annual rate, an action similar to that in 1966 but not as drastic. The result was a "credit squeeze" and a sharp rise in interest rates. Then, according to the critics, came the usual policy reversal in 1970 with the money supply growth rate raised to 5.2 percent. The critics point to a similar experience in 1974–75. The growth of the money supply was reduced in 1973 and then reduced even more sharply in the latter half of 1974. From mid-1974 to early 1975 it grew by less than 2 percent. Federal Reserve policy in 1974 was apparently primarily directed at meeting an interest rate target and to do this called for a severe restriction in the rate of growth of the money supply.[13] To the critics this restrictive money supply growth at least deepened or perhaps even caused the drastic business downturn that started at the end of 1973 and ran to early 1975. The rate of monetary growth accelerated in early 1975 and from then to early 1976 the money supply increased about 6 percent.

The critics who argue that Federal Reserve policy has so often been destabilizing rather than stabilizing find the Federal Reserve au-

[12] M. Friedman, "The Role of Monetary Policy," in *American Economic Review*, March 1968, p. 16.

[13] For an examination of monetary policy during this period, see W. Poole, "Monetary Policy during the Recession," *Brookings Papers on Economic Activity*, 1, 1975, pp. 123–39.

thorities' reliance on interest rates as a guide to policy to be a major cause of its errors. The difficulty of knowing what changes in interest rates may be resisted in the interest of stabilizing financial markets and which may not be resisted without producing instability in output and prices is part of the problem, but more fundamentally the problem rests on the shortcomings of interest rates themselves as a guide. As we saw earlier, Friedman and other monetarists do not believe that changes in the money supply must necessarily be transmitted through changes in interest rates. (We are not concerned here with those changes in the money supply that involve changes in the public's wealth.) Real assets are held to be closer substitutes for money than financial assets. An increase in the money supply can lead to an increase in the demand for real assets and an increase in the rate of production of real assets to the level at which all of the additional money is absorbed in balances held to mediate the greater volume of transactions associated with that level of production. The public need not hold more money than before as a substitute for other financial assets. This view rules out an interest-elastic demand for money and establishes the close link between money and the income level that is the hallmark of the quantity theory.

Regulating the Money Supply by Rule

From this monetarist view of the transmission process, it follows in straightforward fashion that the Federal Reserve should have as its guide to monetary policy the growth of the money supply. If this is granted, the next question is, What is the appropriate rate of growth for the money supply? How do we regulate the

Friedman and those monetarists who support his position answer this question by simply saying that no attempt should be made to use monetary policy in a countercyclical fashion. The lag between the time that action is taken by the monetary authorities and the time the effects of that action are felt on the income level is believed to be both long and variable.[14] Monetary policy can be used in a countercyclical way only if the monetary authorities know the length of the lag between the execution of any policy and the impact of that policy on aggregate demand and income.[15] Then it will be at least potentially possible to conduct an effective countercyclical policy by basing policy actions on a forecast of what conditions will be when the action taken today takes hold. However, in the present state of knowledge, forecasts are frequently far off the mark, and the danger of such error increases with the length of the forecast. If the lag in monetary policy is long, this then is a compelling argument against trying to pursue a countercyclical monetary policy.

Beyond this is also the variability of the lag. Even if accurate forecasting of future economic conditions were possible, simply knowing what the average lag has been in the past is not sufficient information on which to base today's policy, if there has been considerable variability around that average. If the lag may vary from a few months to a few years, the fact that there is no way to predict the length of the lag in any particular case means that an effective countercyclical monetary policy is simply beyond

[14]In comparing turning points in the cycles of the rate of growth of the money stock with the turning points in business cycles as dated by the National Bureau of Economic Research, Friedman and Schwartz found that the peak in the business [illegible]
17.6 months after the [illegible]

reach. This has long been the position of Friedman and some others. However, there is a good deal of difference among monetary economists as to the length of the lag and its variability and so their judgments differ as to whether or not an effective countercyclical monetary policy is possible.[16]

While everyone recognizes that some lag exists and that the lags will not always be the same length, many argue that the lag is neither so long nor so variable that a policy in which the Federal Reserve at least "leans against the wind" will not be better than one in which it stands straight up whichever way the wind is blowing. In reply, Friedman has written as follows:

> We seldom in fact know which way the economic wind is blowing until several months after the event, yet to be effective, we need to know which way the wind is going to be blowing when the measures we take now will be effective, itself a variable date that may be a half-year or a year or two years from now. Leaning against next year's wind is hardly an easy task in the present state of meteorology.[17]

To Friedman these uncertainties rule out any possibility of a successful countercyclical monetary policy. He would have the Federal Reserve abandon its attempts to employ monetary policy in this way and instead follow a rule of expanding the money supply at a fixed rate of about 4 percent per year in line with the long-run rate of growth of the economy, no matter which way the wind was blowing in any particular year. In his judgment, to do other than this is more likely to contribute to instability than to help in achieving stability.[18]

Until the last few years the Federal Reserve had persistently done other than this, its actions bringing about very rapid growth of the money supply sometimes and very slow growth or even actual decreases in the money supply at other times. M_1, the narrowly defined money supply, increased at an annual rate of 5.5 percent from June 1973 to June 1974, at a 1.4-percent rate from June 1974 to January 1975, and then at an 8.9-percent rate from January 1975 to July 1975. Much greater swings in the growth rate occurred during 1971–72. From January to July of 1971, the money supply grew at an 11.6-percent annual rate and then from July to December of 1971 at only an 0.8-percent rate. Over the following year—December 1971 to December 1972—the rate was 8.3 percent. The extraordinarily large swings in the rate of growth of the money supply during 1971–72 may have exceeded what was intended.[19] However, a

---

three lags, the first of which is the "inside" lag, or the lapse of time between the moment at which there is a need for a change of policy and the moment at which the central bank acts. Inside lag is sometimes divided into "recognition" lag and "action" lag, but the time between recognition of the need for action and the taking of action is so short, relatively speaking, that the inside lag in effect becomes the recognition lag. The second lag is the "intermediate" lag, or the lapse of time between the moment at which action is taken and the moment at which the economy finds itself faced with changed money supply and interest rates. The third lag is the "outside" lag noted in the text, or the lapse of time between a change in the cost and availability of credit and the effect thereof on aggregate spending, income, and output. (See T. Mayer, *Monetary Policy in the United States*, Random House, 1968, Ch. 6, and the references given there for further reading.)

[16]See T. Mayer, *Monetary Policy in the United States*, Ch. 6, for a discussion of this subject and p. 183 for a listing of pertinent literature. See also M. J. Hamburger, "The Lag in the Effect of Monetary Policy: A Survey of Recent Literature," in *Monthly Review*, Federal Reserve Bank of New York, December 1971, pp. 289–98.

[17]M. Friedman, *A Program for Monetary Stability*, Fordham University Press, 1960, p. 93.

[18]There have been attempts to measure the results that would have followed if a rule had been adhered to instead of the discretion actually exercised by the Federal Reserve authorities. See L. R. McPheters and M. B. Redman, "Rule, Semi-Rule, and Discretion during Two Decades of Monetary Policy," in *Quarterly Review of Economics and Business*, Spring 1975, pp. 53–64; F. Modigliani, "Some Empirical Tests of Monetary Management and of Rules vs. Discretion," in *Journal of Political Economy*, April 1964, pp. 211–45; and M. Bronfenbrenner, "Statistical Tests of Rival Monetary Rules," in *Journal of Political Economy*, February 1961, pp. 1–14. See also J. A. Richardson, "Monetary Rules and Optimal Monetary Policy," in *Nebraska Journal of Economics and Business*, Autumn 1975, pp. 45–63.

[19]The actual changes in the rate of growth of the money supply do not always indicate the exact intent of the Federal

rapid rate of monetary expansion in early 1971 was still the policy because the monetary authorities were quite concerned with what a lower rate of expansion would do to interest rates. There was fear that "sharp increases in long-term rates at this juncture might have adverse consequences for spending . . . and might thus pose a threat to the economic recovery under way." Interest rates were given emphasis in determining policy, and the growth rate of the money supply was adjusted as needed to prevent undesired rises in interest rates. The latter part of the year saw a decline in interest rates and a reversal of policy toward a much slower rate of growth of the money supply. This experience suggests that, in choosing a compromise between interest rates and the money supply as guides, the authorities at that time placed major emphasis on interest rates.

Implicit in such emphasis is belief in the Keynesian transmission mechanism and a doubt over the monetarist transmission mechanism. As the Federal Reserve authorities viewed the mechanism, the changes they bring about in the money supply affect spending through interest rates. If they want to avoid the contractionary effect on spending that a rise in interest rates would produce, they must move against such a rise by expanding the money supply as needed. Because they saw the mechanism essentially in this way, they were unwilling to follow the monetarist prescription of allowing interest rates to find whatever level they would under a policy of a fixed rate of growth of the money supply. Thus, a situation like that of early 1971 in which interest rates were tending to rise rapidly while the economy was in the first stage of a recovery from a recession is one in which they believed they had to respond by accelerating the rate of growth of the money supply. In so doing, they chose the policy of leaning against the wind.[20]

This choice of policy suggests a rejection of the argument of the more extreme monetarists that the monetary authorities would not face the question of whether to lean against the wind if only they would adopt the rule of expanding the money supply at a fixed rate and allow interest rates to move as they will under this rule. These monetarists believe that what stirs up the wind in the first place is primarily the variations in the rate of growth of the money supply. Thus, had it not been for the policy of sharply reducing the rate of growth of the money supply during 1969 as the authorities leaned against the inflationary winds, there would not have been the recession of 1970, and therefore no recovery with its rising interest rates. The monetarists grant that instability in the economy may arise from sources other than the mismanagement of the money supply by the Federal Reserve authorities, so that a fixed rate of monetary growth is not submitted as a panacea. However, they still hold that if the Federal Reserve were to adopt such a fixed rate as its policy, it would remove the greatest source of instability now at work in the U.S. economy.

Many economists who are unwilling to accept such a rigid policy still favor a policy that would avoid the wide swings in the rate of growth of the money supply that have repeatedly occurred. They would favor allowing the monetary authorities to lean against the wind but to do so only to a degree consistent with a limited variation in the growth of the money supply, and it is toward this kind of policy that the monetary authorities have moved in recent years.

Reserve authorities. As noted earlier, the Federal Reserve does not have complete control over the money supply and it is not possible to hit precisely a target rate of growth over periods of only a few months. The Record of Policy Actions of the May, June, and July 1971 meetings of the Federal Open Market Committee, which sets policy in this area, suggests that a less rapid rate of growth than that which was occurring was wanted, and the record of the September and October meetings suggests that a faster rate of growth than that being realized was desired. The Record of Policy Actions of the Committee for each meeting is released about a month after the meeting and is published in the *Federal Reserve Bulletin*.

[20]The monetarists maintain that a fixed rate of growth of the money supply does not necessarily mean that interest rates will be subject to larger fluctuations than otherwise. They hold that under certain conditions instability in

Target Growth Ranges for Monetary Aggregates

Through the late sixties and into the seventies, more and more members of the U.S. Congress became convinced of the wisdom of a policy of limiting the swings in the monetary growth rate and in 1975 Congress took direct but very limited action on the subject. From the time Congress created the Federal Reserve System in 1913, it had introduced various changes in the structure and the scope of operations of the system, but it had left the formulation of monetary policy completely at the discretion of the Federal Reserve authorities. However, in 1975 it adopted a "sense of Congress" resolution which amounts to a first small step away from a complete "hands off" position. Among other things, this resolution requested the monetary authorities, namely the Board of Governors and the Federal Open Market Committee, ". . . to maintain long-run growth of the monetary and credit aggregates commensurate with the economy's long-run potential to increase production so as to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates."[21] In accordance with this resolution, the Federal Reserve authorities began the practice of publicly specifying each quarter target growth rates for the year ahead for several monetary aggregates. Also, in accordance with another part of the resolution, the Chairman of the Board of Governors now reports on those targets to the Congress each quarter, alternating between the House and Senate Banking Committees. This part of the resolution established for the first time an arrangement for Congressional surveillance of the monetary policy decisions taken by the Federal Reserve authorities.

The key element in the monetary policy decision is the adoption of a growth-rate target for the money supply with a one year horizon. This is the growth rate which the Federal Reserve authorities believe, on the basis of a study of all available information, both econometric and judgmental, to be consistent with the ultimate goals of high employment, price stability, and the like.[22] The practice has been to work with three measures of the money supply, M_1, M_2, and M_3.[23] The target for each is not a specific rate but a range of rates, a so-called tolerance range, which may be whatever the Federal Reserve authorities deems to be consistent with the loose language of the Congressional resolution. Thus, in July 1976 the authorities set the following ranges for M_1, M_2, and M_3 for the year from the second quarter of 1976 to the second quarter of 1977: $4\frac{1}{2}$–7 percent, $7\frac{1}{2}$–$9\frac{1}{2}$ percent, and 9–11 percent, respectively. As M_1 averaged $302.8 billion in the second quarter of 1976, the target for the second quarter of 1977 was M_1 between $316.4 billion, the result of a $4\frac{1}{2}$-percent growth rate, and $324 billion, the result of a 7-percent growth rate. If M_1 during the second quarter of 1977 were to average out at an amount between these two figures, the M_1 growth-rate objective set about a year earlier

---

interest rates is largely the result of instability in the rate of monetary growth, not an alternative to it, because the latter results in instability of price-level changes and this in turn causes instability in the size of the inflation premium that is part of the interest rate. On this subject, see, for example, W. E. Gibson, "Price-Expectations Effects on Interest Rates," in *Journal of Finance*, March 1970, pp. 19–34, and "Interest rates and Inflationary Expectations: New Evidence," in *American Economic Review*, Dec. 1972, pp. 854–65. See also E. Shapiro, "The Monetary Growth Rate and the Interest Rate: A Diagrammatic Presentation," in *Nebraska Journal of Economics and Business*, Spring 1976, pp. 3–17.

[21]See *First Report on the Conduct of Monetary Policy*, Senate Committee on Banking, Housing, and Urban Affairs, U.S. Congress, June 25, 1975, p. 3. Earlier actions of this nature had not gotten beyond the committee level. See *Standards for Guiding Monetary Action*, Joint Economic Committee, U.S. Congress, June 1968.

[22]For an explanation of a quite technical subject, which is here only sketched, see R. E. Lombra and R. G. Torto, "The Strategy of Monetary Policy," in *Economic Review*, Federal Reserve Bank of Richmond, September/October, 1975, pp. 3–14, and W. Poole, "The Making of Monetary Policy: Description and Analysis," in *New England Economic Review*, Federal Reserve Bank of Boston, March/April, 1975, pp. 21–30.

[23]These are defined in footnote 4 on p. 508.

would have been met. In the same way, one can compute the dollar range within which M_2 and M_3 would have to fall in the second quarter of 1977 to meet the objective set about a year earlier for those two monetary aggregates. What is here described occurs every quarter. Early during the fourth quarter of 1976 the authorities, again on the basis of an evaluation of all information then available, set the ranges for the year from the third quarter of 1976 to the third quarter of 1977.

While the objective is to remain within these percentage ranges over the course of a full year, that does not mean that the monetary growth rate in any month or quarter may not be above or below the longer-run range. There is some reason to believe that sizable monthly or even quarterly deviations from the annual growth-rate ranges are not particularly damaging to the economy as long as the results for the year as a whole fall within the ranges set for that time period. In line with this, the objective is to stay within the longer-run growth ranges without at the same time causing unacceptably large variations in short-term interest rates, and to meet the latter objective may call at times for short-run changes in the monetary growth rates that fall outside the longer-run ranges that have been established.

At monthly meetings, the authorities set growth ranges for the monetary aggregates for two month periods, the month of the meeting and the following month, and also set a range for the Federal funds rate for the same time period. The Federal funds rate is the interest rate that member banks charge each other on overnight loans of excess reserves. It is a rate that moves very closely with any easing or tightening of the overall reserve position of the member banks, and it is a rate whose changes are followed by similar changes in other short-term rates like those on U.S. Treasury bills and commercial paper. In conducting open-market operations, which is the major tool used by the Federal Reserve authorities to influence the reserve positions of the commerical banks, the Federal funds rate, and the growth of the money supply, it may be found that the growth rates of the monetary aggregates can not be held within these specified two-month ranges without going outside the specified range for the Federal funds rate. The issue then becomes the one familiar from earlier discussion: Shall priority be given to interest rates and credit conditions or to the money supply? The record shows that the Federal funds rate has been quite consistently held within the prescribed range but the two-month monetary growth rates have frequently been above or below the prescribed ranges.[24] This would suggest that, in the case of month-to-month policy, stability of interest rates takes priority over stability of monetary growth rates.

However, the question of whether the growth rates of the monetary aggregates are consistent with the ultimate objectives of full employment, price stability, and the like is not answered by the short-run fluctuations in these rates. To the degree that the growth rates of the monetary aggregates are as critical as some say they are, it is the rates over a period of time longer than a month or a quarter which appears to be at issue. For the one-year period, which is a more appropriate interval, the record so far is quite limited because an evaluation can only be made after the year in question is completed. However, for the year from the first quarter of 1975 to the first quarter of 1976, the year from the second quarter of 1975 to the second quarter of 1976, and the year from the third quarter of 1975 to the third quarter of 1976, the record suggests that the Federal Reserve authorities did essentially succeed in meeting their publicly announced annual monetary growth objectives.[25]

While no final conclusion can be drawn from such limited experience, it does provide a basis for believing that the Federal Reserve authorities

[24] See S. G. Hoffman, "Monetary Growth Objectives," in *Monthly Review*, Federal Reserve Bank of Atlanta, December 1976, pp. 177–79 for the record through October 1976.

[25] *Ibid.*, pp. 179–81.

have at least sought to conduct policy in such a way as to achieve monetary growth rates that fall within the selected ranges over the course of a year. However, there are complications in this approach that make the actual record difficult to interpret. Is the achievement of appropriate monetary growth rates being given the priority as a policy objective that the Congressional resolution apparently called for or is this priority being unduly qualified by attention of the Federal Reserve authorities to interest rates and credit conditions? Does the resolution imply narrower ranges for the monetary growth rates and wider ranges for the Federal funds rates, steps that would come closer to the "rule" kind of policy recommended by the monetarists? One of the most firmly established facts in the history of U.S. central banking is the propensity of the Federal Reserve authorities to conduct policy in a way that will smooth out the cyclical increases and decreases in interest rates, and it is not easy to detect to what degree this traditional objective is still being adhered to at the expense of the monetary growth-rate targets. The Federal Reserve authorities have retained complete discretion as to what monetary aggregates they will report on to the Congress and the number of these may be as large or as small as they find convenient. The Congressional resolution requests the Federal Reserve to focus on the aggregates but it does nothing to assure that the Federal Reserve will back away from its long-standing preoccupation with the stability of interest rates. However, a first step has been taken and more may follow. With respect to this first step, one specialist's conclusion is that the "major effect is simply to force the Fed to announce some loose numerical targets and, if it subsequently misses them badly, to discuss why . . . . having to report this information dissipates the strict administrative secrecy that formerly surrounded Fed open-market decisions and leaves Fed officials more vulnerable to outside criticism. . . ."[26]

MONETARY POLICY: CONTROLLING AGGREGATE DEMAND BY CONTROLLING THE MONEY SUPPLY

If the present approach to the making of monetary policy evolves into one that holds the growth rates of the monetary aggregates within a quite narrow range, the monetary authorities will have reached the point at which they exercise relatively little discretion in making policy. The monetarists who favor this approach or an even more rigid one maintain that discretionary monetary policy has all too frequently had a procyclical rather than the intended countercyclical impact on the economy. The basis for this position rests in part, as we have seen, on the existence of a lag, which, according to Friedman, is long and variable. It also rests on the alleged power of changes in the money supply to produce (with a lag) proportional or near proportional changes in aggregate demand and thus in the level of money income. If a change in the rate of growth of the money supply from 4 to 10 percent per year (or from 10 to 4 percent per year) causes the growth of aggregate demand to change from 4 to 10 percent per year (or from 10 to 4 percent per year), a change in money growth of this order would indeed exert a powerful effect. If it turns out that it was an error to change the rate by this amount, the damage could be serious.

But if such a change in the rate of growth of money is ordinarily offset in some part by other changes so that the change in the growth

[26] E. J. Kane, "How Much Do New Congressional Restraints Lessen Federal Reserve Independence?", *Challenge*, November/December 1975, p. 37. This article presents an evaluation of the degree of oversight provided by Congress under the resolution, and its answer to the question raised in the title is "not very much."

rate of aggregate demand is much smaller than the change in the growth rate of the money supply, a change of this order in money growth would be much less powerful. In this case an error in policy would have a much less serious effect than the extent of the change in the money growth rate would suggest. In addition, it follows that the monetarist fear that such large changes in the rate of growth of money, even though quite limited in duration, will lead to great instability in the economy is exaggerated. Although the problem presented by lags may still mean that the monetary authorities will at times err and produce a procyclical instead of the countercyclical effect they intend, the extent of such a destabilizing effect cannot be readily inferred from the change in the rate of growth of the money supply. In other words, one cannot see boom or bust in every change of five percentage points in the money growth rate that lasts longer than a quarter or two, although changes of this magnitude and duration at times seem to create fear of such an outcome among some monetarists. To the degree that offsetting takes place, not only is the total impact of any given percentage change in the money supply that is maintained for any given period of time muted to that same degree, but the effectiveness of monetary policy is also decreased.

The word *effectiveness* has been used with several meanings in this area, but one meaning is the ability or inability of changes in the money supply to produce changes in aggregate demand and the income level, and that is the meaning we will give to the word here. If aggregate demand does not rise or fall at all in response to a rise or fall in the money supply, monetary policy is said to be completely ineffective in this sense. To put it positively, to be effective in recession, monetary policy must clearly be capable of increasing aggregate demand, and to be effective in boom and inflation, it must be capable of restraining aggregate demand.[27]

In this final section of our treatment of monetary policy, we will take a brief look at what economists other than monetarists regard as a kind of major limitation on the effectiveness of monetary policy. This is something we have touched on a number of times earlier—the argument that the linkage between money and aggregate demand and therefore between money and the income level is less tight or more loose than the monetarists maintain. However, we will look at it here in a different context from before by noting some of the institutional and structural features of the economy that operate in a way to offset the effect of changes in the money supply and thus tend to loosen the link between the money supply and the income level. Beyond this, we will note some of the features of the economy that prevent the monetary authorities from compensating for this offsetting by simply pursuing a given policy more intensively. Because the results are not the same in both situations, we first distinguish in general terms between the effectiveness of monetary policy in time of recession and in time of boom and inflation and then go on to consider only the latter case in detail: those features of the economy that impose limitations on the effectiveness of a restrictive monetary policy,

[27]Of course, to be effective in a broader sense requires more than this. In recession the increase in aggregate demand generated by expansionary monetary policy must result in an increase in output and not merely an increase in the price level, i.e., the aggregate supply curve must not for some reason turn out to be perfectly inelastic at a recession level of output. In boom and inflation, the decrease in aggregate demand brought about by contractionary monetary policy must result in a decrease in the price level and not merely a decrease in output, i.e., the aggregate supply curve must not turn out to be perfectly elastic at the existing price level. In this broader sense, it is apparent that expansionary monetary policy is more effective, the more elastic the aggregate supply curve, and contractionary monetary policy is more effective, the more inelastic the aggregate supply curve. If one goes beyond the simple aggregate demand and supply curves, which show only absolute changes in output and the price level, to rates of change of output and rates of change of the price level, this conclusion would be restated as follows. Expansionary monetary policy would be more effective, the greater the increase in the growth rate of output and the smaller the increase in the inflation rate; contractionary monetary policy would be more effective, the greater the decrease in the inflation rate and the smaller the decrease in the growth rate of output.

i.e., on monetary policy in time of boom and inflation.

From the years of the Great Depression into the fifties, economists generally held that monetary policy was less effective in stimulating recovery from a depression than in controlling a boom and inflation. Their reference was to inflations of the demand-pull variety that were actually experienced during these years; monetary policy was recognized to be much less effective against inflations of the cost-push variety. Their reference was also to depression or severe contractions, such as those of 1929–33 and 1937–38, and not to such mild downturns as 1948–49 and 1969–70 in the post-World War II period; monetary policy was recognized to be of real assistance in such mild downturns but virtually powerless in the face of very sharp contractions. If we take the opposite combination of mild downturns and cost-push inflations, it then appears that monetary policy is more effective against the downturns than against the inflations. Because our actual experience since the fifties has often involved just such a combination, it may be said that "the wheel has come full circle, and prevailing opinion has returned to the characteristic 1920's view that monetary policy is probably more effective in checking deflation than in checking inflation."[28]

What is the basis for the conclusion that monetary policy is ineffective in the face of a severe contraction? During such a downturn, the Federal Reserve can, of course, through open-market operations pump more and more reserves into the banks and expand their ability to extend credit to all classes of borrowers. But the mere availability of abundant credit at highly favorable terms is not enough to induce borrowing by business persons and consumers to finance additional spending, *if other conditions influencing spending decisons are so unfavorable as to offset the attractive credit terms.* When the economy is in a sharp downturn and the outlook is bleak, business persons do not seek to borrow on short term to build up inventories. Instead they are anxious to reduce inventories and pay off any loans that were previously secured to finance these inventories. Neither do they seek to borrow on long term to finance expansion of plant and equipment; confronted with a business contraction and reduced sales, most find themselves with excess plant and equipment. In the same way, consumers fearful of unemployment or reduction in their incomes do not borrow to finance purchases of automobiles or other durables, expensive vacations, and the like; instead they seek to reduce whatever debts they have already incurred. There are two clichés commonly heard in this connection: "You can't push on a string" and "You can lead a horse to water, but you can't make him drink." In the face of a sharp downturn, there are very few thirsty horses.

This is not to say that an easy monetary policy in times of a severe contraction will be without beneficial effect; its effect will be largely that of preventing a bad situation from getting worse. A restrictive monetary policy combined with a business downturn would surely aggravate the downturn—the classic example of this was the monetary policy in 1931 that contributed to the deepening of the Great Depression. For contractions in general, a restrictive policy could convert an otherwise reasonably orderly liquidation of excessive inventories into a rout as banks called on business persons to repay the loans that had been extended to finance inventories. Forced inventory liquidation of this sort leads to rapidly falling prices for those goods, business losses, and bankruptcies in a process that tends to spread ever more widely throughout the economy. If credit is readily available on favorable terms, it clearly has a stabilizing effect. By meeting the liquidity requirements of business, it can slow and perhaps reduce the extent of the downturn. In the case of a mild downturn, it may be sufficient in itself to reverse the direction of the downward movement.

In a period of boom and inflation, the effectiveness of monetary policy depends in part on the

[28] H. G. Johnson, "Monetary Theory and Policy," in *American Economic Review*, July 1962, p. 366.

nature of the inflation. Unlike the inflations of 1946–48 and 1950–51, the inflations of 1956–58 and the one that started in 1965 and was still going in 1977 appeared to have cost-push elements at work. To the extent that cost-push predominates, both monetary and fiscal policies have quite limited effectiveness in checking the inflation and the boom of which it is a part. Thus, in Chapter 23 we noted the inability of the restrictive monetary and fiscal policy pursued during 1969–70 to bring under control an inflation in which strong cost-push forces were at work at that time. Whatever effectiveness monetary policy has against inflation, that effectiveness is at a maximum when the inflation is predominantly of a demand-pull type, and it is this type with which we will be concerned here. The problem in this case is one of an overly rapid expansion of aggregate demand, and the question is whether the central bank, through its control over the money supply and the cost of money, can check the overly rapid expansion of aggregate demand.

One way of approaching an answer to this very broad question is by analyzing the limitations on the effectiveness of a restrictive monetary policy. If the monetary authorities err in their forecast of the increase in the velocity of money that will occur during a period of expansion, an inflationary expansion of aggregate demand will occur despite the restrictive action taken by the authorities with respect to the supply of money. The public will have found ways of handling a larger dollar amount of spending per dollar of the money supply than was anticipated by the monetary authorities. Thus, unanticipated changes in the velocity of money appear as a first type of limitation. However, even if the monetary authorities do correctly foresee the increase in velocity that will occur, it does not automatically follow that they can adjust for this by pursuing as restrictive a policy with respect to the money supply as is necessary to offset the increase in velocity of money. This brings us to certain limitations standing in the way of the pursuit of a monetary policy sufficiently restrictive to be effective. If there are such limitations that for various reasons cannot be overcome, the effectiveness of monetary policy is thereby equally limited.

Velocity Changes as a Limitation on the Effectiveness of Restrictive Monetary Policy

By means of its general instruments of monetary management, the Federal Reserve can control expansion of the money supply and, within limits, the cost of money. To support an increase in business activity, there must be either more money, a more rapid rate of turnover of money, i.e., a rise in the velocity of the existing money supply, or both. When credit conditions are tightened and the creation of new money through the banking system is restrained, the policy objective of a slowdown in the rate of expansion of business activity will be realized, unless the public finds ways of thwarting this policy through more effective use of the existing money supply. Therefore, in order for the Federal Reserve authorities to succeed in slowing down an overly rapid expansion, they must make a reasonably correct estimate of the change in velocity that may occur when the impact of the restrictive policy is felt as well as of such other matters as the time lag between the implementation of the policy and the impact of that policy on the pace of business activity. They may then gear their policy accordingly. However, the difficulty is that a small error in the estimate of what the velocity will be can mean a large error in the policy action taken. With a 1976 money supply, M_1, of \$304 billion and GNP of \$1,692 billion, the velocity measured as GNP/M_1 was about 5.6 for the year. With the same money supply, a change in velocity from 5.6 to 5.5 or to 5.7, less than a 2-percent change, would mean a swing in GNP of over \$30 billion one way or the other. In carrying out monetary policy, it follows that a fairly small change in velocity, if not anticipated and taken into account by the policy makers, can turn an

otherwise correct policy into just the opposite.

The importance of velocity in this context is brought out by the fact that the percent change in velocity has rivaled or exceeded the percent change in the money supply in certain periods of business expansion. For the first two years of the expansion from the business cycle trough in 1975-I, GNP increased from $1,446.2 to $1,799.3 billion and the money supply, M_1, increased from $283.0 to $315.0 billion. Velocity was thus 5.1 in 1975-I and 5.7 in 1977-I (at annual rates). Over these two years, GNP increased 24.4 percent, the money supply rose 11.3 percent, and the velocity of money jumped 11.7 percent. In this period the percent increase in velocity was greater than the percent increase in the money supply. The actual increase in the money supply was $32.0 billion; the indicated rise in velocity was the approximate monetary equivalent of another $37.8 billion increase in the money supply. The data reveal similar results for other periods of business expansion, one of the most striking being that from 1949-IV to 1953-II. Here the rise in GNP was 41.6 percent, the rise in the money supply was only 15.9 percent, and the rise in the velocity of money was 23.6 percent. The actual increase in the money supply was $17.3 billion but the indicated rise in velocity turns out to be the approximate monetary equivalent of another $25.6 billion increase in the money supply, or much more than the actual increase. To a lesser degree, such results are found for the periods 1965-II to 1969-IV, 1958-II to 1959-II and 1954-II to 1957-III. With changes in velocity being so important relative to changes in the money supply, it is not an easy task for the Federal Reserve authorities to know what the appropriate change in the money supply should be to bring about any desired degree of tightening in the market. There is considerable room for error in estimating what the actual velocity change may be in any time period.[29]

What are some of the ways the public manages to make more effective use of an existing money supply in times of restrictive monetary policy, causing changes in velocity, and thereby complicating the Federal Reserve's problem of deciding on the appropriate change in the money supply? A brief examination of a few of the ways most discussed in the literature will convey the nature of the problem.[30]

1. Commercial Bank Portfolio Adjustments. Increasing business activity brings more borrowers to the commercial banks. In carrying out its restrictive policy, the central bank takes the actions necessary to keep commercial bank excess reserves below the level that would permit the banks to accommodate fully this expanding demand for loans. Although banks may meet customer demand for loans by selling government securities and lending the proceeds to borrowers, this will not actually increase the total amount of credit extended. As the banks sell securities, their total deposits are drawn down as buyers make payment; an offsetting increase in loans simply restores deposits to their previous level. The total earning assets of banks and the total money supply are left unchanged. However, the replacement of government securities with loans in bank portfolios will, at least to some extent, represent a conversion of *idle deposits* into *active deposits*, even through there is no change in total deposits or in the money supply. Some small part of the de-

[29] This is especially true if, as some claim, successive changes in velocity figures from period to period are essentially uncorrelated or that the time series for velocity is characterized as a simple random walk. This rejects the existence of a "trend" in velocity and thus rejects the notion that one can forecast which way velocity will change by noting the changes needed for it to return to "trend" if it has shown "deviations" therefrom or the changes needed to stay on "trend" if that is where it is. See J. P. Gould and C. R. Nelson, "The Stochastic Structure of the Velocity of Money," in *American Economic Review*, June 1974, pp. 405–18.

[30] See H. S. Ellis, "Limitations of Monetary Policy," in N. H. Jacoby, ed., *United States Monetary Policy*, rev. ed., Praeger, 1964, pp. 195–214, and S. W. Rousseaus, "Velocity Changes and the Effectiveness of Monetary Policy, 1951–57," in *Review of Economics and Statistics*, Feb. 1960, pp. 27–36.

posits given up by purchasers of government securities may have been idle deposits; in contrast, virtually all the deposits secured by borrowers will be active, since the purpose of business borrowing is to finance the purchase of goods, the payment of bills, and the like. The process is thus one in which idle deposits are activated and total spending is increased, even though the central bank has kept the total money supply unchanged. The significance of this process obviously depends on the existence of holdings of idle deposits, and the amount of such holdings in the past decade has been very much less in a relative sense than was the case in the fifties and early sixties. The general availability of highly liquid assets on which one can earn more than a negligible rate of interest has made the cost of holding non-earning idle money deposits high enough to practically eliminate such deposits, but some may exist and therefore the process as described can occur to some degree.

There is an important interest-rate effect in this process. If the central bank's actions prevent an increase in the money supply, the prices of securities will fall and their yields will rise as banks sell them in the market. The result is a general upward pressure on the structure of interest rates. This fall in security prices and rise in yields will, via the process of portfolio adjustment, persuade holders of deposits to switch from holding assets in the perfectly liquid form of deposits to holding assets in the less liquid form of securities. But falling security prices will mean capital losses to banks that sell securities to get funds to make loans. Will the reluctance of banks to take capital losses not rule out this process? The so-called locking-in effect of lower prices of securities would not appear to be controlling in most circumstances. It may be controlling if the rise in interest rates is expected to be short-lived, in which case banks may prefer to keep their government securities rather than sell at a capital loss, since the prospect of not only recouping such losses but showing a profit through higher interest rates on loans is not favorable in a short period of time. On the other hand, if banks expect interest rates to go on rising for some time, they will clearly have a profit incentive (and, what may be more important, the advantage of being able to satisfy loan requests of regular customers) in shifting out of government securities into loans. In such case, the capital loss on the sale of securities can be more than recouped through higher interest rates on loans. Furthermore, once loan demand begins to subside, it may be possible for the banks to switch back into government securities at prices below those at which they were sold, thus securing a financial gain in both directions. The shifting composition of bank portfolios over time shows that banks have indeed behaved in this manner in recent years, shifting out of securities into loans in times of restrictive Federal Reserve policy and building up holdings of government securities in times of easy Federal Reserve policy.

This process, by which the commercial banks in effect bring about some increase in the active money supply by moving into loans, raises the velocity of the total money supply. Hence, despite the tight control exercised by the monetary authority over the total quantity of money, the process described above emerges as one loophole in the effectiveness of general monetary controls.

2. The Role of Financial Intermediaries. Some economists have argued that the rapid growth of financial intermediaries in the post-World War II period is another development that has seriously weakened the traditional monetary controls of the Federal Reserve.[31] These intermediaries,

[31] This was emphasized particularly by J. G. Gurley and E. S. Shaw. See their "Financial Aspects of Economic Development," in *American Economic Review*, Sept. 1955, pp. 515–38, and "Financial Intermediaries and the Saving-Investment Process," in *Journal of Finance*, May 1956, pp. 257–76. For a skeptical view of the extent to which intermediaries have weakened monetary controls, see W. L. Smith, "Financial Intermediaries and Monetary

which include mutual savings banks, insurance companies, savings and loan associations, and mutual funds, receive savings from the public and in turn make loans and buy securities in much the same way that commercial banks do.[32] The essential difference, however, is their lack of demand deposit- or money-creating ability. They are called "intermediaries" for the very reason that they do no more than channel savings into the hands of spenders. To the extent that the funds supplied by these intermediaries are derived from current saving by the public, they are acting, in a sense, strictly as "middlemen." The types of loans they make and the securities they purchase affect the allocation of credit, but the credit provided would be matched dollar for dollar by current saving of the public. They could lend more only if the public saved more and placed these savings with them; they cannot create money for this purpose as commercial banks do in the course of their lending by creating new demand deposits.

The argument, however, is that the operations of these intermediaries somehow weaken monetary policy. How is this supposed to happen? One way is through portfolio adjustments similar to those described above for commercial banks. In times of tightening credit, financial intermediaries may shift from government securities to loans with the same effects on the active money supply that follow equivalent commercial bank actions. A second way is unique to financial intermediaries: In times of tightening credit and rising interest rates, they may raise the interest rates they pay on funds placed with them. Attracted by these higher rates, some holders of idle demand deposits will convert them into interest-bearing obligations of the intermediaries. As the intermediaries extend credit on the basis of funds that have been attracted away from commercial banks, idle funds previously carried in demand deposits (or cash hoards) are activated and velocity increases.

This process may occur in the way described because of the highly liquid nature of the obligations issued by some of these intermediaries. Persons who hold assets in the form of savings deposits or savings and loan shares regard them as approximately equivalent to demand deposits or cash in terms of liquidity. Although it is true that they cannot be used directly to make payment, they may in practice be converted quickly into checking accounts or cash, both of which, of course, can be used to make payment directly. In times of tightening credit, rising interest rates on the obligations of these intermediaries thus present the wealth-holder with increasing incentive to convert temporarily idle demand deposits into obligations of the financial intermediaries, for this involves practically no sacrifice of liquidity. Although economists describe such actions as a conversion from money to near-money, those who convert their assets regard the difference in liquidity as negligible.

3. Other Factors. There are a number of other ways in which the private sector has managed to make more effective use of available bank reserves and the money supply. A few of these are the now widespread participation of banks in the Federal funds market as a means of economizing on reserves, the development by finance companies of improved methods of collecting funds and of borrowing funds from various sources all over the country, the introduction of devices such as magnetic ink character codes on checks and computers to read these codes which greatly reduces the time needed to process checks, the tremendous increase in the use of credit cards, and the once-a-month settlement that is made possible with them. Some of these changes, like the greater use of credit cards, may be more a matter of convenience

---

Controls," in *Quarterly Journal of Economics*, Nov. 1959, pp. 533–53.

[32] With respect to their operations in savings deposits, the commercial banks are also in the category of "intermediaries."

than a matter of economizing on the use of money, but tighter money may also lead to greater reliance upon such devices than would otherwise be the case. Under the pressure of tightening money, our highly complex financial system does seek out and find various ways to obtain greater mileage from the available money supply. Because of the difficulty of estimating the short-run impact of a stream of changes like these on the velocity of money, such changes constitute a significant "slippage" in the effectiveness of monetary policy.

Limitations on More Restrictive Monetary Policy as an Offset to Velocity Increases

This brings us to another question. If the monetary authorities could forecast with reasonable accuracy the increases in velocity that might occur under a tight money policy, could they not achieve the desired contractionary effect on the economy by restricting the growth of the money supply to whatever degree was needed to offset the influence of changes in velocity? If a given decrease in the growth of the money supply fails to provide the desired slowing effect on the economy simply because that decrease is offset by an increase in velocity, this appears to say only that a larger decrease in the growth of the money supply is required to do what a smaller decrease would otherwise do. Since the Federal Reserve authorities have the power to slow the expansion of the money supply as they see fit, it would seem that velocity changes do not matter very much, assuming that the authorities can correctly anticipate velocity changes.[33] The difficulty here is that there appears to be a limit to the restrictiveness of policy. A policy so restrictive as to overcome velocity changes may not only check an unhealthy expansion and inflation but also bring about an undesirable contraction in business activity. There are several reasons why the degree of monetary restraint that would be required to overcome this velocity loophole may not be feasible in practice.

1. Prevention of Instability in Financial Markets. To carry out a monetary policy that is sufficiently restrictive to control an inflationary expansion of aggregate demand may call for such extreme action by the central bank as to destabilize financial markets and invite a sharp decline in business activity. During August of 1966, the Federal Reserve nearly reached this point in its attempt to slow the economy. In the words of the Council of Economic Advisers, "Monetary policy was probably as tight as it could get without risking financial disorder. Any further increase in over-all demand could not have been effectively countered by general monetary policy."[34] A similar situation was confronted as monetary policy became very restrictive in 1969 and again in 1973 and part of 1974.[35]

This destabilization of financial markets may come about through the effect that a highly restrictive Federal Reserve policy has on the portfolios of the commercial banks and other institutional investors. In order to raise funds to meet the demands of their business borrowers during the summer of 1966, the commercial banks dumped huge quantities of state and local

[33]It may even be argued that changes in velocity add to the effectiveness of monetary policy by providing a safety valve. If the Federal Reserve misjudges the situation and tightens too much, the rise in velocity will offset the tightening in part and give the authorities time to adjust their policy to less restriction before great damage is done. See L. S. Ritter, "Income-Velocity and Anti-Inflationary Monetary Policy," in *American Economic Review*, March 1959, pp. 120–29.

[34]*Annual Report of the Council of Economic Advisers*, January 1967, p. 60.

[35]A special type of limitation on restrictive monetary policy was faced in early 1973. Apart from any threat to the stability of financial markets, a restrictive policy then threatened to trigger a wage-push inflation, because labor's acceptance of the existing wage-price control program was based, among other things, on an understanding that interest rates would also be prevented from rising. See E. Shapiro, "The Demand-Pull Cost-Push Trade-Off as a Limitation on Monetary Policy: A Graphic Representation," in *Quarterly Review of Economics & Business*, Summer 1974, pp. 19–27.

government securities into the market in a relatively short period of time. Security prices fell, and interest rates rose to near record levels. Although the point of financial disorder was not quite reached in 1966, the point does exist at which the decline in security prices and the rise in interest rates can lead to widespread pessimism, deterioration of expectations, and wholesale cancellation of business investment projects and even of consumer durable goods purchases—developments that in themselves are sufficient to turn the system from inflation to contraction. To avoid such serious results, it would seem that the central bank must try to achieve the same total restraint through a more gradual and cautious policy. The problem with this alternative is that such a policy, although free of the disequilibrating effects of the fast, vigorous policy, may take so long to accomplish what needs to be done that it becomes ineffective as a practical approach. Thus, a moderately restrictive policy may be all that is permissible if disorder is to be avoided, but a severely restrictive policy may be what is called for if an inflationary expansion is to be checked. The central bank must do whatever it can by pressing as hard as it dares.

2. Support of Treasury Debt-Management Operations. With a gross national debt that crossed the $700 billion mark in 1977 and with over one-half of the marketable portion of that debt in maturities of one year or less, the Treasury is almost continuously in the market exchanging new issues for maturing ones or selling new issues to raise funds to pay off maturing ones. In addition to such refunding operations on existing debt, the Treasury has been in the market every year since 1958 except for 1960 and 1969 to raise new money to finance the deficits of those years. The deficits for the six fiscal years, 1971–76, totaled $177.5 billion and raised the gross national debt by 40 percent. With the Treasury faced with figures like these, if the Federal Reserve should undertake a restrictive policy at the same time that the Treasury is faced with a large refunding or cash-borrowing operation, the Treasury may be unable to sell the securities at the prices and yields expected. Because the Federal Reserve has a special responsibility to the Treasury in connection with the latter's debt-management operations, it may therefore be compelled at times to sacrifice the tight policy it would otherwise pursue in order to enable the Treasury to carry out its financing operations successfully. The Federal Reserve authorities each month issue a monetary policy directive to the manager of the system's open market operations and one part of this directive instructs him to take into account Treasury financing operations in carrying out the monetary policy directive. This, at times, is another restraint on the Federal Reserve's ability to pursue the tight money policy that it feels is required in the interests of stability.[36]

3. Discriminatory Effect of a Tight Monetary Policy. A tight money policy effected through the general instruments of control is said to discriminate against particular sectors of the economy. In particular, it is thought to work against small business persons, because they are poorer credit risks, and against residential construction and some types of state and local

[36]A related problem is the effect of a tight money policy on the interest cost of servicing a national debt that in gross terms at the end of fiscal year 1976 was $620.4 billion of which $376.4 billion was held by private investors (i.e., outside of Federal Reserve banks and U.S. government agencies and trust funds). Interest payments on the gross debt were $34.6 billion for fiscal year 1976 or just under 10 percent of the total of Federal budget outlays for that year. With debt service already making up so significant a part of budget outlays, some argue against reliance on tight monetary policy because it will make that part even more significant. However, this is really an argument more against the way monetary controls work than against their effectiveness. If tight policy is otherwise effective, higher interest cost on the debt may be the price to be paid if we are to achieve the stability such a policy helps provide. This differs from the argument in the text. The issue there is against tight money not because of its effect on the cost of debt servicing but because of the difficulties it could present to orderly debt management. The Federal Reserve cannot avoid this responsibility; and, though it limits the effectiveness of monetary policy, it is in the best interest of overall stabilization policy.

government spending, because they are the most sensitive to changes in credit cost.[37] At the same time, big corporations, with their excellent credit ratings and lesser sensitivity to the cost of credit, may find their requirements fully met.

A tight monetary policy is, of course, designed to restrain the total amount of credit extended in order to restrict aggregate demand. But the pressure resulting from tight money, if felt so unevenly in different sectors, will in itself limit recourse to tighter money as a means of restricting aggregate demand. It is impossible for policy-makers to ignore these differential impacts on various sectors, even if they would like to ignore them in the single-minded pursuit of their objective of restricting aggregate demand. The "injured," particularly those concerned with residential housing, a prime victim of tight money, make themselves heard.

There is little likelihood that the monetary authorities will ignore these various limitations on how far they can go in pursuing a restrictive policy. Earlier in this chapter, we noted the strong emphasis they have traditionally placed on the avoidance of instability in interest rates and credit conditions. From the fact that the economy finds ways of offsetting to some degree the restrictive impact of a slowing in the growth rate of the money supply and the fact that the monetary authorities will stop short of a still more restrictive policy when that policy threatens disruption in financial markets or other adverse effects, we may conclude that we have here an important limitation on the effectiveness of monetary policy.

A countercyclical monetary policy faces the serious problem of the lag—the uncertainty as to how long a period of time will elapse before a change in the money growth rate will make itself felt on aggregate demand and the income level. Were it not for this, according to some monetarists, monetary policy could presumably be employed to produce almost any desired change in aggregate demand by adjusting the dial for the rate of money growth as needed. As the majority of economists see it, even if there were no lag, monetary policy would still not have the effectiveness that monetarists would then attribute to it. Velocity is a variable and one whose changes the Federal Reserve authorities can not predict with great accuracy. Therefore, how much of a change in the money supply is needed to produce any desired change in aggregate demand and the income level also can not be determined with great accuracy. While conceding an important role to monetary policy, to the nonmonetarists there is a great deal more needed for a successful program of economic stabilization than control over the money supply.

A CONCLUDING NOTE

Page 1 of this book began with the observation that, in an area that changes as fast as economics, anything that persists for more than a decade qualifies as "traditional." What we have done in the last two chapters of this book is to consider some of the "traditional" body of thought found under the heading of macroeconomic policy. A key proposition in that body of thought is that activist fiscal and monetary policies will produce a better record of output and employment over time than will result if government stands by passively. There has been no shortage of dissenters to this proposition,

[37]See, for example, G. L. Bach and C. J. Huizenga, "The Differential Effects of Tight Money," in *American Economic Review*, March 1961, pp. 52–80; P. F. McGouldrick and J. E. Petersen, "Monetary Restraint and Borrowing and Capital Spending by State and Local Governments in 1966," in *Federal Reserve Bulletin*, July 1968, pp. 552–81 and December 1968, pp. 953–982; N. W. Bowsher and L. Kalish, "Does Monetary Expansion Discriminate against Housing?" in *Review*, Federal Reserve Bank of St. Louis, June 1968, pp. 5–12; and the papers in *Housing and Monetary Policy*, Federal Reserve Bank of Boston, 1970.

the foremost of whom is Milton Friedman, and the nature of and basis for their opposition has been noted in our coverage. It may be added that their dissent has gone on uninterruptedly for so many years that it is itself something traditional in the area of macroeconomic policy. In closing this discussion of macroeconomic policy, it is appropriate to note a new line of dissent to the traditional thought that has come on the scene in just the last few years, but has not yet made its way into the textbooks. In that short time it has created quite a stir in the theory of macroeconomic policy, but it will be some years in the future before any kind of judgment can be made on the correctness or incorrectness of this line of attack.

This attack on the effectiveness of activist fiscal and monetary policy is based on the application of the concept of *rational expectations* to the area of macroeconomic policy. The idea of rational expectations is, in short, that households and firms form their expectations of the future magnitudes of economic variables like the price level, GNP and disposable personal income by using all the information available to them. This includes whatever information they have on what government fiscal and monetary policies will be in the future, because a rational person recognizes that these policies will tend to affect the values of economic variables. In contrast, the conventional approach to the explanation of the way households and firms formulate their expectations, so-called *adaptive expectations*, makes the expected magnitude of these economic variables equal to a weighted average of their present and past values. Therefore, the expected values so determined do not allow for the influence of all the information that rational households and firms take into account in formulating their expectations.

The upshot of this difference is the following. In broad terms, because the public's current spending, saving, investing, and other economic decisions are affected by their expectations of what is going to happen in the uncertain future and because their expectations of this depend in part on their expectations of what macroeconomic policy is going to be, the public's current decisions are what they are, in part, because of what it expects macroeconomic policy will be. What follows in turn is that any policy action that is widely expected to occur, e.g., a cut in taxes, will have little effect on the public's behavior because it will already have been taken into account and acted on by the public. However, any policy action that is unexpected will for this reason cause the public to change its current behavior. Thus, the conclusion is that the policy action will be effective in modifying the public's current behavior in the way intended by the policy makers only if it comes as a surprise to the public. An analogy is provided by the price of a share of stock. That price on any date already reflects all known information that investors regard to be relevant; the only thing that will cause a change in the price of the stock is the appearance of information that was not already taken into account or information that appears as a surprise.

As we saw earlier in this chapter, the Federal Reserve authorities since 1975 have each quarter publicly announced monetary growth targets for the year ahead. Based on this and expected values of other variables like the federal government deficit, the public forms expectations of the rate of inflation. The terms of wage agreements, the level of interest rates, and other variables are then adjusted to reflect the expected rate of inflation. However, the argument is that, because of such adjustments, the policy actions when carried out will not be followed by changes in the level of employment and output, because the public responded to the indicated rate of monetary growth before the steps were taken to produce that rate of monetary growth. On the other hand, if the public has come to expect a certain rate of monetary growth and the monetary authorities without warning boost that rate well above the expected rate, the policy will have the effect of raising employment and output. Whatever the expected rate of inflation had been before, there

will be a higher rate as businesses react to the unexpected actions of the monetary authorities by raising the prices of their output. Because wage rates will lag under these conditions, the reduction in real wages and the rise in profits will mean some increase in employment and output. But the key to this result is the element of surprise, and a countercyclical monetary policy whose success rests on surprise, or some would say trickery, is not one that can be effective on a continuing basis.

According to rational expectations, a similar impediment to the effectiveness of countercyclical fiscal policy is faced. Because the public has learned from experience that there will be tax cuts and/or increases in government spending in the event of an economic downturn or slowdown, signs of such a development give rise to expectations of such countercyclical action. The public acts on these expectations, and the timing of their actions need not provide the desired stabilizing effect. Although it may be an extreme illustration, the investment tax credit has been used to show that not only does business behavior guided by rational expectations rob this device of any countercyclical effect but actually turns it into one with a procyclical effect. The investment tax credit, which has been used off and on since 1962, permits firms to take as a credit against their income taxes a percentage of their outlays for new capital equipment. It appeared to have worked well when it was first used in 1962 because it came as a surprise. However, after a couple of rounds in which businesses found that it is taken off as the economy recovers and put back in when the economy slows, it ceases to work. Worse than this, this particular device becomes perverse. The current investment decisions of business people depend in part on their expectations of what macroeconomic policy will be, and they come to expect that the investment credit will be reintroduced whenever the economy turns down. This leads to postponement of investment spending any time the economy shows signs of a downturn as people seek to take advantage of the investment tax credit, which is expected to be brought in with a downturn. What may not actually have been the beginning of a downturn can be turned into a downturn as the public acts on its expectations of what government policy will be. The downturn comes, Congress provides business with the expected investment tax credit, and investment rises sharply. It looks as if the credit produces its intended countercyclical effect, but the opposite is actually the case. Its effect may be perverse: It starts or accelerates a downturn and then it does the same in the other direction. Instead of being countercyclical, this device turns out to be procyclical, a result that follows from the expectation that government will try to use it in a countercyclical way.

This may be sufficient to bring out the basis for the major tenet of the rational expectations argument: that activist macroeconomic policy cannot systematically succeed in affecting the economy's employment and output levels. To affect these levels, the authorities must alter policy in a way unexpected by the public, but a continuing policy of this kind must lead to uncertainty by the public as to what policy will be. The public's decisions will be influenced by its expectations of what government policy will be, however uncertain these expectations, but this in turn means that the policy makers have reached the point at which they have no way of knowing whether any particular action they take will come as a surprise to the public and thus be effective or whether that action has already been taken into account by the public and will thus be ineffective. Without being able to know what public reaction will follow from any policy step it takes, the policy makers are bound to be in a quandary as to what the appropriate policy at any time should be. Policymaking becomes dice-shooting.

The attack of the "rationalists" is in its early stages and so far has achieved a degree of success, but the contention that rational expectations makes activist macroeconomic policy ineffective is not yet and may never become a widely accepted one. More empirical evidence is

needed and that evidence, whether in support or in opposition or some of both, will be produced in the years ahead. In any event, we have come a long way since the fifties when the effectiveness of countercyclical policy was accepted with little question by the economics profession, apart from a few notable exceptions. It is not beyond the realm of possibility that a decade or two from now the few exceptions will be those who believe in its effectiveness.

appendix

appendix

National Income Accounting

The purpose of this appendix is to explain the major features of the official national income accounting framework developed for the U.S. economy by the Department Commerce.[1] However, to better understand the complexities of that complete framework, it is approached in steps through a series of hypothetical economies. We begin in Section I of this appendix by considering the simplest of all possible economies, one far removed from the actual economy of the United States or any other country. The accounting framework for this economy is made up only of relationships among business firms and households and excludes relationships of these two groups with either the government or other economies. As we saw in Chapter 2, the GNP identity for such a two-sector economy is $C + S = \text{GNP} = C + I$, but here we want to look behind the identity at matters that were not covered in the brief presentation of Chapter 2. In Section II of this appendix, government is admitted into the economy to obtain a hypothetical three-sector economy, and again we look more fully than in Chapter 2 at what lies behind the three-sector GNP identity: $C + S + T = \text{GNP} = C + I + G$. Similarly, in Section III relationships between each of these sectors and other economies are admitted to give us a hypothetical four-sector economy. As we saw in Chapter 2, the GNP identity for the four-sector economy is $C + S + T + R_{pf} = \text{GNP} = C + I + G + (X - M)$, and again we will be looking behind this identity and into details that were entirely bypassed in Chapter 2.

In the accounting framework for the four-sector economy of Section III, we will have a framework that, apart from a number of minor items, is essentially the same as the official one

[1] Attention is here given to national income accounting only. We will not enter into other systems of accounting such as flow of funds, input-output tables, and national balance sheets. For more on national income accounting and treatments of these other accounting systems, see J. W. Kendrick, *Economic Accounts and Their Uses*, McGraw-Hill, 1972; S. Rosen, *National Income and Other Social Accounts*, Holt, Rinehart and Winston, 1972; and W. I. Abraham, *National Income and Economic Accounting*, Prentice-Hall, 1969. For a comparison of the national income accounting systems of the major countries, see M. Yanovsky, *Social Accounting Systems*, Aldine, 1965, and W. Beckerman, *International Comparisons of Real Incomes*, Development Center of the Organization for Economic Cooperation and Development, Paris, 1966. For the system adopted by the United Nations for the international reporting of comparable national accounting data, see *A System of National Accounts*, Department of Economic and Social Affairs, United Nations, New York, 1975.

for the U.S. economy. To simplify the transition from the frameworks for the hypothetical economies of Sections I, II, and III to that for the actual economy, which will be presented in Section IV, many of the definitions and classifications now used by the Department of Commerce will be introduced into the hypothetical cases at appropriate points. For the same reason, on the many occasions in which there are two or more possible procedures for handling a particular item (e.g., to include or exclude interest paid by government from national income), the procedure followed in the accounts for the hypothetical economies will be that followed in the official accounts. All that will remain, then, when we reach Section IV, will be to fit into the framework a number of minor items that were deliberately omitted in order to concentrate on the major items and relationships. These omissions in no sense make the earlier frameworks incorrect but only incomplete, if we take the official framework to be the "complete" one. In Section IV these omissions will be made good.[2] Section V of this appendix looks into some problems in the national income accounts, including an explanation of several items in the accounts, and the final section, Section VI, touches on the relationship between national product and national welfare.

I
NATIONAL INCOME ACCOUNTING: TWO-SECTOR ECONOMY

For the two-sector hypothetical economy with which we start, we can define a fundamental accounting identity that will apply in this economy in any time period:

$$\text{National income} = \text{Net national product} = \text{Expenditures on net national product}$$

or

$$\text{National income} + \text{Depreciation} = \text{Gross national product} = \text{Expenditures on gross national product}$$

in which depreciation is assumed to measure the wear and tear on the economy's capital stock during the period.

The net version of the identity states that the dollar value of the economy's net production of final goods and services, or its net national product, for any time period is identical with expenditures on those final goods and services and with national income earned in the production of those final goods and services. National

[2]Since our purpose in this appendix is to build the framework and describe the essentials of the methodology employed by the Department of Commerce, we may note here the official publications that provide a detailed treatment of material either lightly covered or excluded from this appendix.

The principal results of the latest periodic comprehensive revision of the national income and product accounts have been published in the *Survey of Current Business* for January 1976 in two parts. This report contains a brief analysis of the various definitional and classificational revisions and presents the basic income and product series for 1929–74 as modified by these changes. The changes are the first that have been made since the last overhaul of the U.S. national accounts whose results were reported in the *Survey of Current Business* for August 1965. The most important change, the preparation of estimates of "economic" capital consumption allowances, will be noted in Section IV of this appendix. Most of the definitions underlying the U.S. income and product accounts remain as described in *National Income, 1954 edition* and *U.S. Income and Output, 1958*, U.S. Department of Commerce.

The detailed national income and product statistics for 1929–74 reflecting the definitional and classificational changes introduced in 1976 have been published in *The National Income and Product Accounts of the United States, 1929–74, Statistical Tables, A Supplement to the Survey of Current Business, 1977*. These series are brought up to date each year in the July issue of the *Survey of Current Business*.

income is the sum of wages, rent, interest, and profits, or the sum of the earnings of the factors of production for the time period. For every dollar of final product, there is, on one side, a dollar of income earned in producing final product and, on the other side, a dollar of expenditures on final product.

The measurement of net national product in terms of the economy's expenditures on final product yields a dollar figure based on the prices at which goods are sold in the market. The figure is obtained by adding up all the units of each final good and service sold, weighting each unit at the price at which it is sold. The measurement of net national product in terms of national income, or earnings of the factors of production, yields the same dollar figure in this hypothetical economy. Conceptually, it is a figure that would be secured by adding up all the units of each factor service supplied, weighting each unit at the price paid for, or the income earned by, that unit. The former approach is sometimes referred to as the value of net national product *measured at market prices*, and the latter as the value of net national product *measured at factor cost*. The former approaches final output in terms of the physical product itself and so carries the label of *national product*; the latter approaches final output in terms of the factor incomes that are earned in turning out the physical product and so carries the label of *national income*. The important point is that the value of final product is the same in the present hypothetical economy, whether approached from the sales side as final output times output prices or from the income side as factor input times factor prices.

The gross version of the identity states that the dollar value of the economy's gross production of goods and services, or its gross national product (GNP), for any time period is identical with expenditures on those goods and services and with the sum of national income and depreciation. The gross version thus differs from the net in that it includes as part of product an amount of output excluded from the net version. Specifically, it includes, among other things, the economy's total output of new structures and producers' durable equipment produced during the period, while the net version deducts from this total an allowance for the amount of the total stock of structures and producers' durable equipment that was used up in producing that period's total output. This amount of product is taken to be measured by the depreciation allowances of business for the period, and so it is this amount that in effect is subtracted from both sides of the gross version of the identity to produce the net version.

—CCA

On the income side of the gross version there is now a division of the gross income flow into the portion that constitutes factor earnings, or national income, and the portion that is, in effect, set aside in the depreciation reserves of business. An identity between expenditures on product and income is found in both versions, the essential difference being that the gross version involves a broader definition of product and, correspondingly, a broader definition of income, one in which national income is only a part of a now larger income flow.

In a two-sector economy, final product, whether defined in net or gross terms, is the sum of product produced by or originating in the business sector and product produced by or originating in the household sector. In practice, the amount originating in the household sector is relatively small, and for simplicity we will assume here that it is zero. Therefore, by measuring the amount of final product originating in the business sector, we will be measuring final product for the two-sector economy as a whole. From the fundamental identities set forth earlier, we may accordingly arrive at a figure for the economy's net national product by measuring those expenditures on goods and services purchased from business that are purchases of net final product or, alternatively, by measuring the amount of national income that originates in the business sector. Similarly, we may arrive at a figure for the economy's gross national product by measuring those expendi-

tures on goods and services purchased from business that are purchases of gross national product or, alternatively, by measuring the sum of the amount of national income and depreciation allowances that originate in the business sector.

Most of the basic data required for this purpose are found in the ordinary profit-and-loss statements of business firms. It is through a combination, reorganization, and consolidation of these data that a business income-and-product account emerges, showing, on one side, the amount of product originating in business and, on the other side, the amount of income originating in business. Such an account will be constructed below.

PRODUCT ORIGINATING IN THE SINGLE FIRM

We begin with a highly simplified Sales and Cost-of-Goods-Sold statement for a single firm, as shown in Table 1. The right-hand side of the account indicates the firm's total sales receipts of $225 for a particular time period. The left-hand side of the account indicates the allocation of these receipts. The costs chargeable against total sales are $205, so the residual is a profit of $20. If Cost of Goods Sold had been $245 rather than $205, the residual would have been a loss of $20 rather than a profit of this amount. The profit or loss figure is whatever amount is required to make the allocations side balance with the receipts side.

Suppose we had statements like that of Table 1 for each firm in our two-sector economy. At first glance, it might seem that we could get a meaningful total for the economy's product simply by adding up the value of the goods sold or the sales receipts of all firms. There are several reasons why we cannot proceed in this way, the most obvious of which is the double counting involved. Table 1 shows that goods and services purchased from other firms at a cost of $80 were used up in the course of producing goods and services sold for $225. The firms from which Firm A purchased the $80 in goods and services show this same $80 among their sales. To include this same $80 as part of the output of Firm A clearly involves double counting of this $80. Beyond this, there may be triple and quadruple counting if behind these sales of

TABLE 1

FIRM A
Sales and Cost-of-Goods-Sold Statement
January 1-December 31, 1977

Cost of goods sold		$205	Sales	$225
Purchases of goods and services from other firms	$80			
Wages and salaries	75			
Interest	20			
Depreciation	30			
Profit		20		
Total current costs plus profit		$225	Total sales receipts	$225

$80 by other firms to Firm A are purchases by these other firms from still other firms of goods and services that are used up in producing the goods and services sold for $80. The very same good could be counted over and over as it passed through successive firms on its way to a final purchaser. If this were done, we would be counting and recounting *intermediate product*. The total then derived by adding up the value of goods sold by all firms would in no sense be a meaningful total for the economy's output. To avoid any double counting, we must remove from the total of the value of goods sold by all firms the amount of goods that are intermediate in the production process. What then remains is a meaningful total for the economy's output that is known as *final product*.

As covered in Chapter 2, if we subtract from a firm's sales the value of the goods and services purchased from other firms that were used up in producing the goods sold, the remainder is the *value added* by that firm. To arrive at a total for the economy's output that can properly be called final product, one must count each firm's contribution to that total by the amount of value added by that firm. Thus, Firm A's contribution to production is $145 or $225 − $80. Subtracting $80 from $225 on the right-hand or receipts side gives us a figure that measures the value of product originating in the firm. In the same way, subtracting the same amount of intermediate product on the left-hand or allocations side leaves the items whose sum equals value added. In the present simplified statement, what remains on the allocations side as value added are the various factor earnings, including the residual figure for profit, plus one nonfactor charge against receipts—depreciation.

This approach to the measurement of product originating in the firm gives correct results as long as there is no change in the firm's inventories during the time period. Table 1 includes as costs only the current costs incurred in producing the goods and services *sold* during the period. However, the $80 of purchases from firms is not necessarily the total of purchases from other firms; it is the value of purchased goods and services that were used up in producing the goods and services sold during the period. Similarly, the figure for wages and salaries in Table 1 is not necessarily the total of wages and salaries paid; it is the amount chargeable against the sales for the period. The simplified statement does not show purchases from other firms, wages and salaries paid, and other factor charges that were incurred in the production of goods that were not sold but were instead added to inventories. In order to allow for this part of the firm's output, we must work

TABLE 2

FIRM A
Production Statement
January 1-December 31, 1977

Purchases of goods and services from other firms	$ 90	Sales	$225
Wages and salaries	80	Change in inventories	15
Interest	20		
Depreciation	30		
Profit	20		
Charges against total value of production	$240	Total value of production	$240

with an expanded statement for the firm that includes the change in inventories and the total of purchases from other firms and the total of factor payments.

Table 2 shows the modifications in the statement of Table 1 that result if the firm shows an increase in inventories of $15 during the period. To show the total production of the firm, this change in inventories is added to the total for sales given in Table 1. Total value of the firm's production is thus $240, or the sum of sales of $225 and the increase in inventories of $15. The same additional $15 is included on the allocations side, here assumed to be made up of an additional $10 of purchases from other firms and an additional $5 in wages and salaries.[3]

Although the value of Firm A's production now appears to be $15 higher than the figure earlier derived from Table 1, its actual contribution to the economy's final product is only $5 greater than the earlier figure—against the $15 added to inventories, we now have the additional purchases of $10.[4] Since the product side now includes not only sales but also the change in inventories, to arrive at the value of the firm's contribution to production on the product side we must subtract not only purchased goods used up in producing the output *sold* but the total of *all* purchased goods, including those that are reflected in the change in inventories. Thus, in Table 2, subtracting the full $90 of purchases from both sides of the account is necessary to avoid double counting. What remains on the allocations side is value added of $150—a figure $5 higher than before, which reflects the additional $5 in wages and salaries paid by the firm. The same total of $150 is found on the product side after the corresponding subtraction from the total there.

The figure of $150 we have now arrived at may be described as the *gross* amount of national product originating with this firm for the period. It is called gross because no deduction was made for depreciation, the figure for which may be interpreted as the value of the amount of capital goods used up by the firm in producing its output for the period. In deriving the $150 figure from Table 2, we subtracted from both sides the total amount of purchases from other firms, but these purchases were restricted to services and to those goods that are treated as expenses by the firm. Such purchases may be described as purchases of *noncapital goods*. The firm also purchases goods of a durable type, described as *capital goods*, which are not *expensed* during the period but are *depreciated* over a number of periods. Because their full purchase price is not charged against the production of the accounting period during which they are acquired, such purchases do not appear in either of the statements above. What does appear in both is an entry for depreciation, which is the portion of the total cost of the firm's existing stock of capital goods that is charged against the particular period's production.

We earlier saw that the firm's purchases of noncapital goods and services had to be subtracted from the product side of its account to avoid double counting. During the course of the firm's productive activities, it also uses up capital goods, whether or not it buys any during the period, and a correct reckoning of the firm's contribution to the economy's product seems to call for a deduction of this amount as well. From this viewpoint, $150 (or $240 − $90) as the measure of the firm's contribution to the economy's product involves an overstatement

[3]Depending on the firm's inventory-valuation procedure, the same physical addition of goods to inventories might be valued at a higher or lower figure. For example, if these goods had been valued at $16 instead of $15, assuming as before that an additional $10 was spent for purchases of goods and services and an additional $5 for wages and salaries, the remaining $1 would appear as an additional $1 of profits. The other items on the allocations side—interest and depreciation, in this case—are in the nature of fixed costs and would remain unchanged during the period, whatever the total on the product side happened to be.

[4]Not all the $10 of purchases was necessarily used up during the period in producing whatever the firm produces. Some part could remain in the form in which it was purchased or in the form of goods in process. Whatever its form, however, it is valued on some basis and included in the figure for change in inventories.

of $30, which is the amount of the firm's depreciation for the period. If we deducted this $30, our final figure would be $120, which may be described as the *net* amount of national product originating with this firm for the period. Both the gross and net figures have their place, and both will appear in the accounting frameworks to be developed.

PRODUCT ORIGINATING IN THE BUSINESS SECTOR

In the actual procedure of estimating the amount of gross and net national product originating in business firms, national income accountants do not proceed on an individual-firm basis as was done here for the sake of illustration. They begin with combined data for all firms (or for large groups of firms, such as corporate and noncorporate) and, through a process of consolidation, estimate figures for gross national product and net national product originating in all firms (or, again, for large groups of firms).[5]

Account 1 of Table 3 is such a combined statement, each entry in which is the sum of the values for the corresponding entry in all the individual firm statements. All the individual statements, in the simple form of Table 2, will include the same entries except for the entry for profits. In the combined statement, the net income of all noncorporate firms appears as "proprietors' income," and the net income of corporate firms as "corporate profits," broken down into the portion paid out in dividends and the portion retained.

Total sales in the combined account are broken down into sales to households and sales to business, with the latter broken down into sales that (from the viewpoint of the buyer) are purchases of (1) capital goods or (2) services and noncapital goods. Since each firm shows only purchases of noncapital goods on its own statement, it necessarily follows in the combined statement that purchases of $300 on the allocations side must be identical with $300 in sales to business of noncapital goods on the product side. Furthermore, since these purchases are composed of goods and services either used up by the firms in producing the goods and services sold or else incorporated in the change in inventories, the amount involved, $300, must be subtracted from both sides to eliminate double counting. This adjustment yields a consolidated account, Account 2 of Table 3, whose total of $425 on the product side is equal to expenditures on GNP and whose total on the allocations side, or what now may also be called the income side, is the sum of national income and depreciation.

Gross national product is defined to include the total purchases by business on capital account. However, as was noted in examining the production statement for Firm A by itself, a definition that includes in final product all the capital goods produced during the period is a definition that involves some double counting. Just as purchases of noncapital goods had to be deducted from both sides to avoid double counting, so, it may be argued, an amount of purchases of capital goods equal to the value of capital goods used up during the period must be deducted for the same reason. If the $45 of depreciation for all firms is taken as a measure of the value of capital goods used up during the period, this amount may be subtracted from both sides, reducing sales of capital goods to business from a gross figure of $60 to a net

[5]Recall that we are assuming in this section that all product and income originates in the business sector, so we may simply refer, in what follows, to gross national product instead of gross business product or gross product originating in business, and similarly for net national product and national income.

TABLE 3

ALL FIRMS COMBINED
Production Statement
(Account 1)

Purchases of goods and services		$300	Sales to households	$355
Wages and salaries		320	Sales to business of capital goods	60
Interest		35	Sales to business of services and noncapital goods	300
Depreciation		45	Change in inventories	10
Proprietors' income		5		
Corporate profits		20		
Dividends	$15			
Undistributed profits	5			
Charges against total value of production		$725	Total value of production	$725

ALL FIRMS CONSOLIDATED
Business Income and Product Account–Gross Basis
(Account 2)

Wages and salaries		$320	Sales to households = personal consumption expenditures	$355
Interest		35	Sales to business of capital goods = gross fixed investment	60
Proprietor's income		5	Change in inventories	10
Corporate profits		20		
Dividends	$15			
Undistributed profits	5			
National income originating		$380		
Depreciation		45		
Charges against business gross product		$425	Business gross product	$425

ALL FIRMS CONSOLIDATED
Business Income and Product Account–Net Basis
(Account 3)

Wages and salaries		$320	Sales to households = personal consumption expenditures	$355
Interest		35	Sales to business of capital goods less depreciation = net fixed investment	15
Proprietors' income		5	Change in inventories	10
Corporate profits		20		
Dividends	$15			
Undistributed profits	5			
National income originating (= Net national product originating)		$380		
Charges against business net product		$380	Business net product	$380

figure of $15. This adjustment yields Account 3 of Table 3, a consolidated account whose total of $380 on the product side is equal to net national product and whose same total on the allocation or income side is the sum of factor earnings or national income.

This brings us back to the fundamental accounting identities with which we began. In a two-sector economy, GNP may be measured by the sum of national income and depreciation or by the sum of expenditures on goods and services that are included in the GNP. In Account 2 of Table 3, depreciation is $45, and the sum of factor earnings or national income is $380, making a total of $425 on the income side of the account. On the product side of the account, what is a sale from the seller's viewpoint is, of course, an expenditure from the buyer's viewpoint. Thus, sales to households of $355 may also be described as personal consumption expenditures of this amount, and sales to business on capital account of $60 may also be described as fixed investment expenditures by business or simply fixed investment of this amount. The increase in inventories of $10 may also be viewed as a type of expenditure in which the business sector purchases from itself goods that are added to its inventories. Thus, the sum of expenditures on goods and services included in gross national product is $355 + $60 + $10; or GNP as measured by expenditures is $425. The same total for GNP results from adding the items on either side of this account. For the fundamental identity for GNP with which we started, we have the following figures:

National income								+	Depreciation	=	Gross national product	=	Expenditures on gross national product				
Wages and salaries	+	Interest	+	Proprietors' income	+	Corporate profits							Personal consumption expenditures	+	Gross fixed investment	+	Change in inventories
$320	+	$35	+	$5	+	$20		+	$45	=	$425	=	$355	+	$60	+	$10

For the two-sector economy, the identity in net terms states that national income is identical with expenditures on net national product. To move from the gross to the net national product identity simply involves eliminating the entry for depreciation on the income side and subtracting an equal amount from gross fixed investment on the product side. This leaves national income of $380 on the income side and personal consumption expenditures of $355, net fixed investment of $15, and the change in inventories of $10 on the product side, or $380 as the total of expenditures on net national product. For the fundamental identity for net national product with which we started, we thus have the following figures:

National income							=	Net national product	=	Expenditures on net national product				
Wages and salaries	+	Interest	+	Proprietors' income	+	Corporate profits				Personal consumption expenditures	+	Net fixed investment	+	Change in inventories
$320	+	$35	+	$5	+	$20	=	$380	=	$355	+	$15	+	$10

Corresponding to each of the two definitions of product, we thus find an identity between the amount of expenditures on that product and the amount of income generated in the production of that product. In the case of GNP, not all the income is made up of earnings of the factors, or national income; a part of that income flow is, in effect, set aside by business in depreciation reserves. Of the income flow that constitutes earnings of the factors, the larger part becomes income receipts of persons who have provided labor and property services to business. It may be noted that these income receipts in turn are the primary sources for financing the expenditures that appear on the product side of the business-sector account.

SECTOR ACCOUNTS IN THE TWO-SECTOR ECONOMY

As a final step, the complete set of accounting relationships in this simple two-sector system may be brought out by expanding Table 3 to include a sector account for households and a saving and investment account for the economy as a whole. Account 1 of Table 4 is basically the same as Account 2 of Table 3. Account 2 of Table 4 is a household-sector account that shows on the right-hand side the income receipts of households, the total for which is known as *personal income*, and on the left-hand side the allocation of these receipts. In the present economy, which omits government and foreign transactions (and also assumes that no income originates in households), personal income originates in its entirety from payments made by business. All the wages and salaries, interest, and proprietors' income generated in the business sector passes on to households to become personal income. However, only that part of corporate profits paid out as dividends passes on to become personal income; the remainder, or undistributed corporate profits, is retained by business and is a part of gross business saving. The receipts side of the household-sector account accordingly shows personal income of $375, an amount $5 less than national income. Of this amount, persons devoted $355 to personal consumption expenditures as shown on the allocation side. (This is the same entry that appears on the product side of Account 1, the business-sector account.) The amount of personal income not devoted to personal consumption expenditures is personal saving, in the present case $20.

Account 3 of Table 4 presents a fundamental accounting relationship: *the identity between the economy's gross saving and its gross investment*. Account 3 is not a sector account—it does not indicate anything like receipts and allocations or product and income on its two sides. Rather, it summarizes the economy's transactions on capital account for the period and shows that, by definition, for every dollar of investment in the economy there is a dollar of saving, and vice versa. *Investment* may be defined simply as the amount of the economy's *product* that is not consumed during the period, while the amount that is consumed is taken to be measured by personal consumption expenditures during the period. In the illustration, GNP is $425, and personal consumption expenditures are $355. Unconsumed output, or gross investment, is then $70, the sum of gross fixed investment and the change in inventories.

Saving is defined in a manner parallel to the definition of investment—it is the amount of *income* not devoted to personal consumption expenditures. Since the size of the gross income flow is by definition equal to the size of the gross national product, saving as the amount of gross national *income* not consumed is identical with investment as the amount of gross national *product* not consumed. Investment was seen to

be \$70 for the period, so saving must also be \$70 for the period. It can be shown to be so, as follows: Of the gross income flow of \$425, two parts, depreciation and undistributed profits, are the only ones retained by business. The sum of these two items, \$50, equals *gross business saving.* The balance of the gross income flow of \$375 passes on to households to become personal income. Of the \$375 of personal income, \$355 is devoted to personal consumption expenditures. The remainder of \$20 is the saving of the household sector, or what is called *personal saving.* The sum of gross business saving of \$50 and personal saving of \$20 is accordingly \$70, an amount equal to gross investment.

It is important to see that this identity between saving and investment holds under any

TABLE 4

BUSINESS INCOME AND PRODUCT ACCOUNT
(Account 1)

Wages and salaries		\$320	Personal consumption expenditures		\$355
Interest		35	Gross private domestic investment		70
Proprietors' income		5	Gross fixed investment	\$60	
Corporate profits		20	Change in inventories	10	
Dividends	\$15				
Undistributed profits	5				
National income originating		\$380			
Depreciation		45			
Charges against business gross product		\$425	Business gross product		\$425

PERSONAL INCOME AND OUTLAY ACCOUNT
(Account 2)

Personal consumption expenditures	\$355	Wages and salaries	\$320
Personal saving	20	Interest	35
		Proprietors' income	5
		Dividends	15
Personal outlay and saving	\$375	Personal income	\$375

GROSS SAVING AND INVESTMENT ACCOUNT
(Account 3)

Gross private domestic investment		\$70	Personal saving	\$20
Gross fixed investment	\$60		Undistributed profits	5
Change in inventories	10		Depreciation	45
Gross investment		\$70	Gross saving	\$70

and all circumstances. Suppose, for example, that the economy's output of final goods and services remains at $425 for the period, but that only $350 of the corresponding income flow is devoted to personal consumption expenditures. This means that $75 of the income flow is devoted to saving. It also means, perforce, that $75 of the economy's $425 worth of product is not consumed, since that much has not been matched by consumption expenditure. And, by definition, the amount of the economy's output that is not consumed is its investment. Investment is therefore $75 and equal to saving. If we varied the illustration in the other direction, with $360 of the income flow of $425 being devoted to consumption expenditures, by the same reasoning we find that investment will be $65, equal to saving of $65.

The derivation of the identity between gross saving and gross investment may be summarized in the steps shown in Table 5, which uses the appropriate figures from Table 4. Canceling personal consumption expenditures from the income side leaves the amount of gross income not devoted to consumption expenditures, or the amount of gross saving; canceling personal consumption expenditures from the product side leaves the amount of gross product not consumed, or the amount of gross investment:

$$\underbrace{\underset{20}{\text{Personal saving}} + \underset{\$5}{\text{Undistributed profits}} + \underset{\$45}{\text{Depreciation}}}_{\substack{\text{Gross saving} \\ \$70}} = \underbrace{\underset{\$60}{\text{Gross fixed investment}} + \underset{\$10}{\text{Change in inventories}}}_{\substack{\text{Gross investment} \\ \$70}}$$

The Gross Saving and Investment Account of Table 4 shows the same breakdown of gross saving and investment derived above. The totals on the two sides of the account or, what is the same thing, the two sides of the identity, will always be identical; the economy's gross saving and gross investment are by definition the same dollar magnitudes.

It is now an easy matter to convert Table 4 from a gross to a net basis. For Account 1, we can use Account 3 of Table 3, instead of Account 2 of that table. This gives us the smaller *net* product of business, and the correspondingly smaller *net* income flow—both of which are less than the gross amounts by the $45 of depreciation. The household-sector account is unaffected, since the removal of $45 for depreciation from the gross income flow is the removal of a form of business saving, a part of the income flow that does not pass on to households. The saving and investment account will now be a *net* rather than a *gross* amount. On the saving side will be the personal saving of $20 and the *net* saving of business of $5, equal to undistributed profits, giving a total for personal and net business saving of $25. On the investment side will be net fixed investment of $15 (gross fixed investment of $60 less the $45 deduction for depreciation) plus the change in inventories of $10 as before. The total of net investment, $25, is identical with the total of net saving, $25.

Viewed in terms of the identity between saving and investment, we derive the *net* identity from the *gross* identity by simply deducting an amount equal to depreciation from both sides. On the investment side, we may first separate gross investment into net fixed investment and replacement investment. Then the deduction on the product side amounts to the exclusion of the amount spent by business for replacement of capital goods used up, the figure for depreciation being taken as the measure of the capital goods used up. We start with the gross identity:

$$\underbrace{\underset{\$20}{\text{Personal saving}} + \underset{\$5}{\text{Undistributed profits}} + \underset{\$45}{\text{Depreciation}}}_{\text{Gross saving } \$70} = \underbrace{\underset{\$45}{\text{Replacement investment}} + \underset{\$15}{\text{Net fixed investment}} + \underset{\$10}{\text{Change in inventories}}}_{\text{Gross investment } \$70}$$

After subtracting the amount of depreciation from the saving side and its equivalent, replacement investment, from the investment side, we have:

$$\underbrace{\underset{\$20}{\text{Personal saving}} + \underset{\$5}{\text{Undistributed profits}}}_{\text{Net saving } \$25} = \underbrace{\underset{\$15}{\text{Net fixed investment}} + \underset{\$10}{\text{Change in inventories}}}_{\text{Net investment } \$25}$$

TABLE 5

		National income			+	Depreciation	=	Gross national product	=			Expenditures on gross national product		
		$380			+	$45	=	$425	=			$425		
Personal income			+	Undistributed profits	+	Depreciation	=	Gross national product	=	Personal consumption expenditures	+	Gross fixed investment	+	Change in inventories
$375			+	$5	+	$45	=	$425	=	$355	+	$60	+	$10
Personal consumption expenditures	+	Personal saving	+	Undistributed profits	+	Depreciation	=	Gross national product	=	Personal consumption expenditures	+	Gross fixed investment	+	Change in inventories
$355	+	$20	+	$5	+	$45	=	$425	=	$355	+	$60	+	$10

CIRCULAR FLOW IN THE TWO-SECTOR ECONOMY

The production and sale of final product and the generation of income that accompanies these activities are processes that take place on a continuous, day-to-day basis. Table 4 presents figures that are accumulations of amounts over some accounting period, say a year. Thus, the figure for wages and salaries ($320) is a total for the year, but part of this total was paid out in the first week, part in the last week, and parts in the other fifty weeks of the year. Personal saving of $20 is a total for the year that is likewise composed of the amounts of saving during each shorter period into which the year may be divided. Similarly, for every item in Table 4, the figure is the cumulated total or the flow of this item over a time period here assumed to be one year.

As shown by the household-sector account, certain of these flows are combined to produce the aggregate flow known as *personal income.* Similarly, as shown by the business-sector account, certain of these flows are combined to produce the aggregate flow known as *business gross product.* The relationships among the aggregate flows of gross product, net product, national income, and personal income may be derived from a study of Table 4, but the flow nature of these same relationships may be seen more clearly by recasting the table into the circular-flow diagram of Figure 1. The diagrammatic presentation immediately focuses attention on a basic feature of the economy—the circular nature of the flow of payments from firms to households and of expenditures from households to firms. Thus, the upper loop of the figure shows a physical flow of productive services from households in exchange for a monetary flow of income from business in payment for these services; the lower loop, at the same time, shows a physical flow of consumer goods and services from firms in exchange for a monetary flow of expenditure from households. These two flows may also be viewed as one circular flow in real terms and one in monetary terms. The former is a clockwise flow of real productive services from households to firms and real goods and services from firms to households; the latter is a counterclockwise flow of monetary income from firms to households and monetary expenditure from households to firms.

Recall our assumption that all product is produced in the business sector, making gross product, net product, and national income originating in business the same as the economy's gross national product, net national product, and national income. Thus we see these three aggregates on the right-hand side of Figure 1, which shows firms. National income is equal to factor earnings, but the amount of these earnings paid out by firms is less than total factor earnings by the amount of undistributed corporate profits. National income is $380, but factor payments to persons are $375, the difference of $5 being the amount of national income retained by firms. In the present case, the only source of personal income is from factor payments by business, so we find the flow of factor payments from business reappearing on the left-hand, or households, side as an equal amount of personal income. Personal income is allocated to personal consumption expenditures, $355, and personal saving, $20, as shown in the lower loop of the figure.

In the present economy, national income and net national product are the same amount, the amount of factor income generated in producing the goods and services that constitute net national product being equal to the amount of expenditures on these goods and services. These expenditures are personal consumption expenditures and net business investment. With

consumption expenditures given as $355 and net national product as $380, net investment, as the part of net national product not consumed, is equal to $25. The identity between net investment and net saving is seen in the figure by combining personal saving of $20 coming out of households and net business saving of $5 for a total of $25, or an amount equal to net investment. The small loop on the right includes the $5 of undistributed corporate profits (net business saving) and suggests the sense in which this amount of saving that occurs within firms joins the flow of saving coming from households to make up a total of net saving by definition equal to net investment.

Gross national product exceeds net national product by the amount of depreciation. Like undistributed corporate profits, depreciation is an amount of the gross income flow retained by business and appears together with undistributed corporate profits as the total of business saving in the small loop on the right-hand side of the figure. Given that gross investment is the part of GNP not consumed, and given further that GNP is $425 and that the amount of this consumed is $355, gross investment appears as $70. The identity between gross saving and gross investment is seen by combining the $20 of personal saving coming out of households with the gross business saving of $50 shown in the small loop for a total of $70, or an amount equal to gross investment.

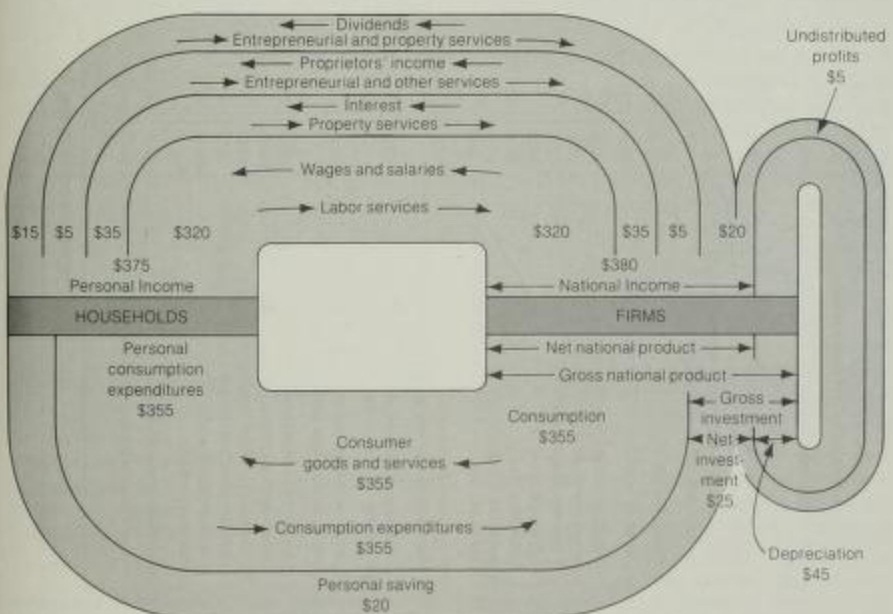


FIGURE 1
Circular flow in the two-sector economy

The system of accounts developed in Table 4 and the representation of that system in Figure 1 have been limited to an economy made up of business firms and households only. In the next section the system of accounts and the circular-flow diagram of this section will be expanded to incorporate the government sector and to reflect the changes in the system that result from the inclusion of government transactions.

II NATIONAL INCOME ACCOUNTING: THREE-SECTOR ECONOMY

We first considered a two-sector economy in order to permit an analysis of the basic product and income, saving and investment relationships uncomplicated by government transactions. But an economy without government spending or taxation is a far cry from any real economy. In the United States today, every area of the private sector of the economy feels the impact of government transactions. Over the past decade, about 23 percent of the GNP has consisted of government purchases of goods and services, and direct taxes on persons and corporations have amounted to about 17 percent of national income. With the addition of government transactions to our framework, all the identities developed earlier will have to be modified. None, however, need be rejected.

The structure of Table 6 parallels that of Table 4 but involves the addition of new entries in the various accounts as well as of entirely new accounts.[6] One of the new accounts, Account 3, is a consolidated statement of government receipts and expenditures, consolidated for the thousands of government units—federal, state, and local—that have the power to tax, spend, and incur debt. Intergovernmental transactions are canceled out in the consolidation, so the totals shown reflect only the position of the government sector relative to each of the other sectors of the economy. Any entry that appears as a receipt in one consolidated account must be an allocation of another (or in some cases the same) account, and vice versa. For each item included in government receipts, there is another account (or opposite side of the same account) in which that particular government receipt appears as an allocation in the form of a tax payment. Similarly, for each item included in government expenditures, there is another account (or opposite side of the same account) in which that particular government expenditure appears as a receipt. If government expenditures are greater than tax receipts, there is a deficit; if they are less, there is a surplus. Unlike the other items in the government account, which lead to the business or household account, the deficit or surplus appears as government dissaving or saving in the gross saving and investment account (Account 5), which brings together all the economy's saving and investment.[7]

[6] Because of these additions, the dollar figures for the entries in Table 6 unavoidably differ from those for the same entries in Table 4. No direct comparison of dollar figures in these two tables is therefore possible.

[7] If all government units consolidated show a deficit, the amount of the deficit or the amount borrowed may be shown in the government account on the receipts side as a plus item or on the expenditures side as a minus item. It will here be considered as a minus item on the expenditures side. A surplus accordingly will be considered as a plus item on the expenditures side.

GOVERNMENT RECEIPTS

All government receipts may be treated as tax receipts if we include among tax receipts social insurance contributions and incidental nontax revenue such as fines and license fees.[8] Tax receipts are classified as of four types: personal taxes, corporate profits taxes, social insurance contributions, and indirect business taxes (including fines and license fees).

Each of these four types of taxes may be traced to the other sectors of the economy for which they are payments or allocations. Personal taxes ($24) are an allocation of the household sector; corporate profits taxes ($10) and indirect taxes ($25) are allocations of the business sector; and social insurance contributions ($10) are allocations of the business, government, and household sectors.[9] Wages and salaries paid by business, government, and household employers (total of $141) are shown inclusive of employee contributions but exclusive of employer contributions for social insurance. This total amount of wages and salaries is treated as if it passed on to households in its entirety. Then the amount of employee social insurance contributions ($4) paid on their wages and salaries is subtracted from total receipts of households ($204), the remainder being "personal income" ($200).

GOVERNMENT EXPENDITURES

While many classifications of government expenditures are possible,[10] at this point the appropriate classification for national income purposes is simply (a) those expenditures for which government receives either goods or services (i.e., those that are matched by productive activity) and (b) those expenditures for which government receives neither goods nor services (i.e., those that are not matched by productive activity). In Table 6, government expenditures matched by productive activity are composed of goods and services purchased from business ($30) and services of labor purchased directly from government's own employees ($20). The balance of government expenditures ($20)—those not matched by productive activity—is composed of transfer payments ($13), subsidies less current surplus

[8] Government enterprises engaged in production, such as local water supply systems, are treated in the U.S. income and product accounts in much the same way as are private profit-making firms. Their receipts are primarily from sales of services rather than taxes, and for this reason they are included as a part of the business sector of the economy rather than a part of the government sector.

[9] Government receipts from social-insurance contributions in our hypothetical economy total $10: $4 is collected from the business sector as employer contributions, $1 from the government sector as employer contributions (the amount of $1 appears as both a receipt and an expenditure of government), $1 from the household sector as employer contributions, and the remaining $4 from households as employee contributions. Households and certain institutions included in the household sector employ persons and pay the employers' share of social-insurance contributions on wages and salaries of these employees. The figure here of $1 is disproportionate with wages and salaries paid but is used to avoid figures of less than $1.

[10] The most familiar is by function: defense, general government, interest, education, etc., in the case of the federal government.

TABLE 6
Accounts for three-sector economy

BUSINESS SECTOR (Account 1)

Wages and salaries		$118	Personal consumption expenditures (or sales to households)		$155
Employer contributions for social insurance		4	Government purchases of goods and services (or sales to government)		30
Rent		5	Gross investment		40
Interest		8	Change in inventories	$ 5	
Proprietors' income		25	Purchases of capital goods	35	
Corporate profits		20			
Profits taxes	$10				
Dividends	6				
Undistributed profits	4				
National income originating with business		$180			
Business transfer payments		1			
Indirect taxes		25			
Current surplus of government enterprises		1			
Less Subsidies		4			
Net national product originating with business		$203			
Capital consumption allowances (depreciation)		22			
Gross national product originating with business		$225	Gross national product of business		$225

HOUSEHOLD SECTOR (Account 2)

Personal taxes				$ 24	Wages and salaries		$141
Personal outlays				161	From business	$118	
Personal consumption expenditures			$160		From government	19	
					From households	4	
Purchases from business		$155			Interest		13
Purchases from households		5			From business	$8	
Wages and salaries	$4				From government	4	
Employer contributions for social insurance	1				From households	1	
					Rental income of persons		5
Interest paid by consumers	1				Income of unincorporated business		25
Personal saving				15	Dividends		6
					Transfer payments		14
					From business	$ 1	
					From government	13	
					Less Personal contributions for social insurance		4
Personal taxes, outlays, and saving				$200	Personal income		$200

TABLE 6 (continued)

GOVERNMENT SECTOR (Account 3)

Purchases of goods and services			$50	Personal taxes			$24
From business		$30		Corporate profits taxes			10
From employees		20		Indirect taxes			25
Wages and salaries	$19			Contributions for social insurance			10
Social insurance contributions	1			Employer contributions		$6	
Transfer payments			13	From business	$4		
Net interest paid			4	From government	1		
Subsidies less current surplus of				From households	1		
government enterprises (4 − 1)			3	Personal contributions		4	
Deficit			−1				
Government expenditures and deficit			$69	Government receipts			$69

NATIONAL INCOME AND PRODUCT (Account 4)

Wages and salaries		$141	Gross national product originating with	
From business	$118		Business sector	$225
From government	19		Government sector	20
From households	4		Household sector	5
Employer contributions for social insurance		6		
Proprietors' income		25		
Corporate profits		20		
Profits taxes	$10			
Dividends	6			
Undistributed profits	4			
Rent		5		
Interest		8		
National income		$205		
Business transfer payments		1		
Indirect business taxes		25		
Current surplus of government enterprises		1		
Less Subsidies		4		
Net national product		$228		
Capital consumption allowances (depreciation)		22		
Gross national product		$250	Gross national product	$250

GROSS SAVING AND INVESTMENT (Account 5)

Gross investment		$40	Personal saving		$15
Change in inventories	$ 5		Business saving		26
Purchases of capital goods	35		Capital consumption allowances (depreciation)	$22	
			Undistributed profits	4	
			Government saving		−1
Gross investment		$40	Gross saving		$40

of government enterprises ($3),[11] and interest payments ($4).[12]

Government expenditures for goods and services cover goods from paper clips to aircraft carriers and services from those of a clerk to those of the President. In the United States, the bulk of these expenditures consists of payments to business firms for the goods and services purchased from them; the balance consists of wages and salaries paid directly to government employees. Just as business firms decide whether to "make or buy," within limits government decides which goods and services it will produce on its own and which ones it will purchase from business firms. For example, if a government agency needs 10,000 copies of a government report, the job can be handled either in the government's printing plant or by an outside printing firm. In either case, the cost to government will be included as part of government purchases of goods and services. If the reports are printed by the government, part of the total expenditure will be in the form of wages and salaries to government employees and part in the form of purchases of paper, ink, and other supplies from business firms. If they are printed by a private firm, the total government expenditure will be in the form of purchases from business firms. Therefore, in Account 3 of Table 6, of the total of government purchases of all kinds of goods and services ($50), part may be traced to the receipts side of the business sector (sales to government of $30), and part may be traced to the receipts side of the household sector (wages and salaries from government of $19 or, if employer contributions for social insurance on these wages and salaries are included, $20).

We have noted that government expenditures, other than for goods and services, consist of transfer payments ($13), subsidies less current surplus of government enterprises ($3), and net interest payments ($4). Transfer payments to persons[13] include benefits paid under various social-insurance programs (old age and survivors, unemployment, and railroad retirement insurance) and other programs such as direct relief and military pensions. Subsidies are monetary grants by government to business. Net interest is simply interest paid on public debt less interest received by government. All of these expenditures have one characteristic in common: The government receives no goods or services in exchange for them—no productive activity matches the expenditure.

Government transfer and interest payments are treated as receipts of the household sector. Although in reality a large portion of the interest payments on public debt is received by banks, insurance companies, and other firms, these interest payments are considered, in effect, to pass through business firms and to become, in their entirety, receipts of the household sector. Accordingly, the total of net interest paid by government appears directly as a receipt of this sector.

Sales to government appear on the receipts side of the consolidated business account, since the payment is received in return for goods and services produced for government.

[11] Subsidies of $4 less current surplus of government enterprises of $1. In the official accounts, only the difference is available, due to the fact that many subsidy programs are interwoven with the operations of government enterprises. An explanation of current surplus of government enterprises is provided later in this section.

[12] A question arises as to whether interest paid by government is a payment for a productive service. Net interest paid by business is plainly factor income to the lenders of funds. It is income received by them in exchange for a productive service (namely, the provision of funds to business) that contributes to additional production of goods and services and thereby produces additional income to business out of which the interest is paid. The major part of net interest paid by government is paid on debt incurred primarily to finance wars and the aftermath of wars. Since such debt cannot, in the ordinary sense of the word, be regarded as debt incurred for productive purposes, interest paid on such debt similarly cannot be regarded as a payment matched by productive activity. In the United States income and product statistics, interest paid by government is not viewed as a return to productive activity and is treated as the equivalent of a transfer payment. An exception to this rule had been government interest paid to foreigners, but this was changed in the 1976 revisions. This is explained in Section III of this appendix.

[13] There are also foreign transfer payments made by government; these will be considered in the four-sector economy of the following section.

Government subsidies to business might also seem to belong on the receipts side of the business account. However, there is no productive activity to match the receipt of such subsidies, and their amount would have to be removed before the receipts side could be used as a measure of production originating in business. Therefore, government subsidies appear in Account 1 as a negative item on the allocations side. By appearing in this form, this placement also reflects the fact that the subsidies result in lower prices for the output of subsidized business and/or higher incomes to its owners and employees. That is, the increased receipts that subsidies would seem to represent are, in fact, balanced by lower receipts from sales (due to lower prices) or by larger total factor payments (ordinarily in the form of higher profits or higher wages and salaries) or by a combination of both.

GOVERNMENT PRODUCTION

In the two-sector economy, the question of production by government did not arise. All production originated in business firms.[14] As a result, measuring gross and net national product and national income originating with business firms was the same as measuring gross and net national product and national income for the economy as a whole. If government is now recognized as a producer, the measurement of income and product for the economy as a whole requires that income and product originating with government be added to that originating with business.

That government is a producer is readily seen if we include under government the many business-type agencies, such as publicly owned local transit and water-supply systems, whose costs are covered, at least to a substantial extent, by the sale of goods and services to their customers. For national income accounting purposes, the fact that these agencies are publicly rather than privately owned does not alter the fact that they are essentially business enterprises. Unlike private business enterprises, their survival does not depend on the earning of profits, but the business aspect of their operations is still paramount. These agencies, which are referred to as "government enterprises" in the U.S. income and product accounts, are treated as part of the business sector. Accordingly, the government sector is limited to those government agencies whose services are not or are only incidentally sold in the marketplace and whose expenses are covered almost entirely by taxes.

The government sector that emerges from this division is, for accounting purposes, similar to a mammoth household. Whatever it purchases from the business sector is final product in the same way that whatever the household sector purchases from the business sector is final product. If a city purchasing department buys 10 gross of pencils, these pencils are final product just as they would be if purchased by households. The government sector does not resell what it purchases; therefore, what it purchases is not viewed as intermediate product. However, the same purchase of pencils by a business firm (either a private firm or a government enterprise) is clearly intermediate product, for the pencils will be used up in the course of producing the firms' output, and the cost of pencils will be charged against gross sales in order to arrive at that firm's final product.[15]

[14]In the two-sector economy of Chapter 2, we assumed for simplicity that no production originates in the household sector. Actually, a relatively unimportant amount of production does originate in this sector and will be considered later in this chapter.

[15]Other things being equal, if the firm does not use up during the period all the pencils purchased, the remainder will appear as an increase in inventory and as such is part of the period's final product.

While these purchases by the government sector from the business sector are thus included as part of final product, they are clearly final product *produced* by the business sector. Does the government sector itself (distinguished from government enterprises) produce anything? Or does this sector merely tax and borrow on the one hand and spend on the other without producing anything in the process? What if government taxed and borrowed and spent the funds so raised entirely either on the purchase of goods and services produced by business firms or on transfer payments, subsidies, and interest? The former are goods and services produced by business, not by government. The latter involve no goods or services secured by government and, accordingly, no production on the part of either government or business. However, these two are not the only types of government expenditures; there remain government expenditures in the form of wages and salaries paid to employees. These employees are paid for the labor services they provide to government and thus to the public. The amount that they are paid may be taken as a measure of the amount of production these services represent.

Let us illustrate by returning to the 10,000 copies of a government report that could be printed in a government printing plant or by a private printing firm. In measuring the nation's net national product, the 10,000 copies of this report should be included at some dollar value in the economy's output of final goods and services, regardless of where they were produced. If the report is printed by a private firm, it is included in net national product, at the dollar amount paid by government to the printing firm, as product originating in the business sector. As a purchase by government, it is part of final product; as production of the business sector, it is part of final product originating in the business sector. If it is printed by government in its own printing plant, only part of its value will be made up of paper, ink, and other supplies that were purchased from business. The balance of its value will be the amount of wages and salaries paid by government to its employees for their labor services on this report. Assuming that the purchase price charged by a firm equals the cost to government if the report is produced in its own printing plant, the publication should enter the net national product at the same dollar value in either case. However, in the second case, it will appear in net national product at this dollar amount only if we include as part of net national product government expenditures for wages and salaries paid to its own employees as well as government purchases of materials from business. If we multiply this example a thousandfold, not to include wages and salaries paid by government would result in a large but unreal increase in net national product any time government shifted work from its own employees to those of business firms. Conversely, the reverse shift would result in a large but unreal decrease in net national product. In short, not to include wages and salaries paid by government would be to understate the value of production in any period.

The same reasoning applies to whatever government acquires for its expenditures on goods and services, whether it be 10,000 copies of a report or some other good, or whether it be the provision to the public of the services of policemen and firemen, schoolteachers and research workers, doctors and lawyers, or of any of the hundreds of other services rendered by government employees. The contribution to production per time period of every government employee may be measured by the amount of his wages or salary for that time period. Thus, total production per time period by the government sector may be measured by government's expenditures for the services of its employees per time period. This, in effect, measures production originating in government on a cost basis, and the only cost of production included is labor cost. Otherwise viewed, government production or output is measured in terms of input, and the only input recognized is labor. In principle, then, in the official income and product statistics, the value of government production

is measured by the wages, salaries, and supplements thereto paid by government.

This method of measuring production in the government sector is seen to differ radically from the method described earlier for measuring production in the business sector. The measurement of production in the business sector can be approached conceptually from either the receipts or the allocations side of that sector's account. On the receipts side before consolidation, we found gross sales of goods and services to all sectors of the economy, including interfirm sales or sales within the business sector. After adjusting gross sales to show the amount of noncapital goods and capital goods used up in producing the gross amount of goods and services sold, and after further adjusting the receipts side to show any change in inventories, the remainder on the receipts, or sales, side is net production of final goods and services by the business sector.

A corresponding approach to a measure of government production on the receipts side is ruled out by the fact that government does not sell what it produces. For the business sector, the market prices at which final goods are sold dictate the dollar value at which these goods are to be included in the nation's final product. The government sector, however, does not sell but "gives away" practically all the goods and services it provides to the rest of the economy. It is true, of course, that, apart from a deficit, the services provided by each governmental unit are paid for through taxes; but taxes are not like prices, since there is no direct correspondence between the amount of product in the form of public services that a taxpayer receives and the amount of taxes that he pays. Even if there were a direct correspondence, the taxpayer is not given any real choice about how much he will pay in taxes or how much he will "purchase" in public services. Thus, the compulsory aspect of taxes paid to government, as compared with the volitional aspect of purchases made from business, rules out the possibility of measuring government production through government receipts in the way that production by business is measured through business receipts.[16]

If, then, government production must be measured from the expenditures or allocations side of the account, the measurement is necessarily in terms of factor cost and thus parallels the method followed in measuring business production from the allocations side of the business-sector account. The measurement task is more complex for the government sector, since differences of judgment can and do arise as to what items in government expenditures are to be included as factor cost. For example, as noted earlier, the United States accounts exclude net interest paid by government as a factor cost, whereas the reverse treatment is followed in the official accounts for some other countries. For our purposes, however, in measuring government production ($20 in Table 6), the method followed will be that found in the official accounts: Government production is valued at factor cost, and factor cost is equal to the compensation of government employees.

HOUSEHOLD PRODUCTION

The definition of production within the household sector is the same as that within the government sector.[17] Like the government sector, the household sector purchases goods and services from the business sector and directly from persons employed by the household sector. Again, following the procedure in the official accounts, the only direct purchases of factor services recognized in the household sector are

[16]To illustrate further, assume that government receipts in the form of contributions for social insurance are doubled from $10 to $20 billion per year and are matched by an

purchases of labor services, and the amount of these services is measured by the wages and salaries paid to the sector's employees, including domestic servants of households and the employees of the various institutions and funds included in this sector. The sum of wages and salaries paid to its employees equals national income originating within this sector and is the measure of production within this sector.[18]

From Account 2 of Table 6, we see that the amount of production within this sector is $5. Households pay household employees $4 in wages and salaries, and they pay $1 in employer contributions for social insurance; $4 of these allocations appears as receipts of the same sector under wages and salaries, and the other $1 appears as a receipt of the government sector.

Here as with the government sector, to exclude the direct purchases of labor services by the household sector would be to understate the economy's total final product. For example, a householder may hire a business firm to wash the walls of his house, or he may give this task to a servant. In the former case, the expenditure for this service would appear as a sale to households by the business sector and as such would be part of final product originating in the business sector. If the same task were performed by a household servant (and, say, at the same cost), then unless the wages paid to the servant for this task appeared as factor income of the employee, final product for the economy as a whole would be lower than if the work were performed by a business firm. Since the very same production occurs in both cases, total wages and salaries paid to household employees must be included in final product to reflect the amount of services produced by these employees and the amount of income originating in the household sector.

Although interest paid by consumers is not regarded as a part of national income and product originating in the household sector, it is a part of personal outlays and appears under this heading on the allocations side of the household-sector account. This account shows interest of $1 paid by consumers and interest received of $13 ($8 from business, $4 from government, and $1 from households). The $1 of interest from households on the receipts side is the same amount as the $1 of interest paid on the allocations side. The $8 of interest paid by business is part of national income originating with business and appears in its full amount as

---

equal increase in government transfer payments per year. Despite the great increase in government receipts, this change does not affect government production at all. These dollars merely pass through government and are in no way related to production. In the same way, if an increase in tax receipts is matched by an equal increase in purchases of goods and services from the business sector, there is no change in government production. There may be additional production, but this is production of the business sector. Finally, suppose there is an increase in government receipts that is matched by an equal increase in government payrolls. In this event there is a matching increase in government receipts and government production, taking as the measure of government production its expenditures for wages and salaries. But even though this increase in government production may accompany an increase in government receipts, the two are in no fixed way tied together. With no fixed tie, there is no practical way of measuring government production from the receipts side of the government-sector account.

[17]The household sector is broader than the name suggests. In the official accounts it includes not only households in the ordinary sense of the word but also nongovernmental, nonprofit institutions, such as hospitals, churches, schools, clubs, and nongovernmental pension, health, and welfare funds.

[18]As in the case of interest paid by government, interest paid by consumers is not regarded as reflecting production in the household sector. Although this treatment is more controversial than the same treatment of government-paid interest, it is defended in part by the fact that, unlike interest paid by business, interest paid by consumers cannot be regarded as measuring the contribution of consumer capital to production. As in the case of government, debt resulting in interest payments may be incurred by consumers without a corresponding acquisition of capital assets. As noted above, the bulk of federal interest payments is the result of debt incurred during war, to which no acquisition of productive capital corresponds. Similarly, much of the interest paid by consumers results from loans for purposes other than the acquisition of capital assets. Interest paid by consumers on mortgage debt continues to be regarded as a payment reflecting production. However, the purchase of residential structures by persons is treated in the official accounts as a business transaction, and the interest paid on the mortgage debt that results appears as net interest originating in the business sector. This will be explained in Section V.

a receipt of households. The $4 of interest paid by government is excluded, for reasons already indicated, from national income originating with government but is part of the interest income of households. Thus, in the present three-sector economy, the interest component of national income for the economy as a whole is the $8 of interest that originates with business.

Since income and product originating with the household sector are defined as wages, salaries, and supplements paid by households, total income originating with this sector in Table 6 equals $5. In contrast to income originating in business ($180) and government ($20), the amount originating with households is, by present definition, a relatively small part of the total. But it could be many times larger under another definition—for example, if we included not only the wages and salaries actually paid to employees of this sector but also an imputed amount representing the wages and salaries that would be paid if housewives sold their services to the household instead of providing them free. Under present definitions, a wall-washing job in the home appears as part of gross national product whether it is done by a business firm or by a paid household employee. If it is done by the housewife herself, however, this output is not included as part of gross national product, despite the fact that the product in question is the same in all three cases.

The purpose of the present series of hypothetical economies is not to enter into these questions but to show how an economy's income and product are measured with a given set of definitions of income and product. In Section V we will examine these definitions more closely to see exactly what is included and excluded from the income and product of each sector, and from this we will see what the resulting aggregates, such as national income and gross national product, seem to measure.

PRODUCTION FOR THE ECONOMY AS A WHOLE

We are now in a position to combine the income and product of the government and household sectors with that of the business sector to arrive at the national income and GNP for the three-sector economy as a whole. The accounts of Table 6 are designed to trace the flows of income and product among the three sectors, so that the allocations of each sector are the receipts of other sectors and vice versa. On the basis of the data in these accounts, another set of accounts, Table 7, has been set up to focus attention specifically on the income and product originating with each of the three sectors. A comparison of Tables 6 and 7 will show that Accounts 1 and 4 of Table 7 merely repeat the same accounts of Table 6, for reasons that are explained below.

For the household and government sectors, net and gross national product originating with each sector equal national income originating with each. This follows from the definition of final product in each of these sectors as factor income directly generated by these sectors. In the business sector, however, net and gross national product originating with business exceed factor income originating with business (see Account 1 of Table 7). This occurs because final product is valued at the market prices at which the sector's goods and services are sold, but not all the proceeds from the sale of output at market prices become national income or income of the factors of production. The principal additions to national income originating with the business sector that must be made in order to balance national income and GNP originating with this sector are indirect business taxes and

TABLE 7
Accounts for income and product originating by sector

CONSOLIDATED BUSINESS INCOME AND PRODUCT ACCOUNT (Account 1)

Wages and salaries		$118	Personal consumption expenditures (or sales to households)		$155
Employer contributions for social insurance		4	Government purchases of goods and services (or sales to government)		30
Rent		5	Gross investment		40
Interest		8	Change in inventories	$ 5	
Proprietors' income		25	Purchases of capital goods	35	
Corporate profits		20			
Profits taxes	$10				
Dividends	6				
Undistributed profits	4				
National income originating with business		$180			
Business transfer payments		1			
Indirect taxes		25			
Current surplus of government enterprises		1			
Less Subsidies		4			
Net national product originating with business		$203			
Capital consumption allowances (depreciation)		22			
Gross national product originating with business		$225	Gross national product of business		$225

depreciation allowances, both of which have been considered earlier. Three other adjustments are needed, however, in order to account fully for the difference between national income and gross national product originating with this sector:[19]

1. Business transfer payments must be added.
2. Subsidies must be subtracted.
3. Current surplus of government enterprises must be added.

Business transfer payments, like government transfer payments, are payments in exchange for which the payer receives no goods or services. In the business sector these transfer payments include gifts, prizes, and scholarships given by business to in-

[19] In the official accounts there is a fourth adjustment, one for the *statistical discrepancy*. The Commerce Department estimates the items on the income side and the product side independently and terms the difference in the totals for the two sides the "statistical discrepancy." In practice, the discrepancy is shown on the income side. If the estimate for the income side exceeds the estimate for the product side, the discrepancy appears as a negative item; if the other way around, it appears as a plus item.

This discrepancy, it may be noted, indicates not the *accuracy* of the estimates but their *consistency*. It is possible to have numerous large inaccuracies without a correspondingly large statistical discrepancy, because these inaccuracies may offset one another. For example, a $3 billion underestimate of personal consumption expenditures accompanied by a $3.1 billion underestimate of compensation of employees will in itself only involve a $0.1 billion statistical discrepancy. Of course, in the accounts of the hypothetical economies now before us, no such discrepancy appears.

TABLE 7 (continued)

HOUSEHOLD INCOME AND PRODUCT ACCOUNT (Account 2)

Wages and salaries		$4		
Employer contributions for social insurance		1		
National income originating with households		$5	Net and gross national product of households	$5

GOVERNMENT INCOME AND PRODUCT ACCOUNT (Account 3)

Wages and salaries		$19		
Employer contributions for social insurance		1		
National income originating with government		$20	Net and gross national product of government	$20

NATIONAL INCOME AND PRODUCT ACCOUNT (Account 4)

Wages and salaries		$141	Gross national product originating with	
From business	$188		Business sector	$225
From government	19		Government sector	20
From households	4		Household sector	5
Employer contributions for social insurance		6		
Proprietors' income		25		
Corporate profits		20		
Profits taxes	$10			
Dividends	6			
Undistributed profits	4			
Rent		5		
Interest		8		
National income		$205		
Business transfer payments		1		
Indirect business taxes		25		
Current surplus of government enterprises		1		
Less Subsidies		4		
Net national product		$228		
Capital consumption allowances (depreciation)		22		
Gross national product		$250	Gross national product	$250

dividuals, and also the bad debts of individuals to business. As such, these payments are not factor income for the recipients but receipts for which no productive activity was involved. Although not factor payments, they are as much charges against business product as are wages and salaries or other factor costs. To omit them would produce a dis-

crepancy between gross national product of business on the product side and the charges against gross national product of business on the income side. Yet to include them as factor income would falsely inflate the national income originating in the business sector. Hence, although they are included as a charge against gross national product originating with business, they are excluded from national income originating with that sector. Instead, they appear as an item to be added to national income in going from national income to gross national product originating with the business sector.

Subsidies are similar to transfer payments in that no goods or services are provided to the payer of subsidies—that is, to government.[29] They appear in the business sector as a negative item on the income side. As was noted earlier in the consolidation of the business-sector account, subsidies (which are actually a receipt of the business sector) are subtracted from both sides of the account, so that the total on the receipts side only reflects receipts from production. As a result of the receipt of subsidies, the prices of business output are lower, factor payments originating with business are higher, or some combination of both occurs. Let us assume the result is lower sale prices. Then subsidies as a negative item on the income side offset lower sale prices and the resultant lower sale receipts on the product side. Alternatively, assume the result is higher factor incomes. Then subsidies as a negative item on the income side offset higher factor incomes on that same side, leaving sale receipts on the product side unchanged. On either assumption, if subsidies were not deducted on the income side, a discrepancy would result between the income and product sides of the account.

It will be recalled that government enterprises are included in the official accounts as part of the business sector. *Current surplus of government enterprises* is the difference between the receipts from their sales of final product and certain costs, thereby giving the term "surplus" the flavor of "profits," as that term is used in private business. As computed in the official accounts, however, depreciation, interest, and taxes are not included among the costs of government enterprise, so that the residual "current surplus" is quite different from the related residual "profits" in private business.

Like private business, government enterprises sell goods and services in the marketplace, pay wages and salaries to employees, and buy goods and services from other firms. Consequently, their sales of final product appear as part of GNP originating with the business sector, and their payments of wages and salaries appear as part of national income originating with the business sector. The difference between the value of their final product and wages and salaries paid is called *current surplus,* which may be positive or negative. Since the surplus is not recognized as profit, it is not factor income and therefore is not part of national income. However, since surplus is the balancing item between sales of final product on the product side and charges against this product on the income side, the current surplus must be included somewhere. Accordingly, it is shown as an item to be added to national income to arrive at GNP originating in the business sector.

Account 4 of Table 7 is obtained by adding Accounts 1, 2, and 3, which show GNP and

[29] Not all payments by government that may appear to be subsidies are so classified here. For example, in the case of agricultural support programs, direct cash grants to farmers or losses on nonrecourse loans made by the Commodity Credit Corporation, a government enterprise, are counted as subsidies. However, in the case where government purchases products at artificially high prices as a means of supporting the incomes of the producers of these products, the full amount of such purchases is counted as government purchases of goods and services, despite the fact that an obvious element of subsidy is present.

national income originating in the business, household, and government sectors of the economy, respectively. Thus, on the product side of Account 4 is shown the amount of GNP originating with each of the three sectors, and on the income side the charges against this GNP. These charges consist of the national income originating in all three sectors plus indirect taxes plus business transfer payments minus subsidies plus current surplus of government enterprises plus capital consumption or depreciation allowances of the business sector. Notice that in going from national income for the economy as a whole to its net and gross national product, the required adjustments are all found in the business sector account, or Account 1.

THREE MEASURES OF OUTPUT

As noted in Chapter 2, gross national product, net national product, and national income are the principal aggregate measures of product in the U.S. accounting framework, and we here review the relationships among them for the three-sector economy presently being examined. National income, though it appears on the income side and not on the product side of the national income and product account, is nonetheless a measure of product. It is a measure of product with product valued at "factor cost," or in terms of the factor income earned (though not necessarily received) by the factors of production. As such, it is the narrowest measure. However, the proceeds from the sale of final goods and services at market prices are not matched by factor costs alone but by factor costs plus indirect taxes and business transfer payments and current surplus of government enterprises less subsidies. If we add all these elements, we get the broader measure of final output that is *net national product*. Although net national product is always greater than national income, they are both measures of the same physical output, the difference being in the basis of measurement (i.e., in the valuation of output) and not in the thing being measured (i.e., the output itself). Finally, the proceeds from the sale of final goods and services at market prices can be taken to exclude or include the proceeds from the sale of capital goods for replacement purposes—that is, it can include among the proceeds an amount equal to either net investment or gross investment—which leads to the distinction between *net* national product and *gross* national product. Recognizing gross investment as part of final product yields a larger total on the product side—larger by the excess of gross investment over net investment. A part of the proceeds from the sale of this larger product, exactly equal to the excess of gross over net investment, also appears as a charge on the income side in the form of capital consumption allowances. When capital consumption allowances are added to factor income plus other adjustments (indirect taxes plus business transfer payments plus current surplus of government enterprises less subsidies), the income side of the account exactly balances the product side, as shown in Account 4 of Table 7.

While national income and net national product are both measures of the *same* physical product, gross national product is a measure of a different and *larger* physical product, which includes the amount of capital goods used up during the period's production. If we subtract from gross national product the amount of capital goods used up during the period, we have net national

product, which shows that net and gross national product measure two different physical amounts of output. Viewed in this way, there are only two measures of the economy's output, a *net measure*, which at market price gives us net national product and at factor cost gives us national income, and a *gross measure*, which is gross national product.

The relationships among these several measures of output may be shown as listed in the column on the right.

Gross national product		$250
Less	Capital consumption allowances	22
Equals	Net national product	$228
Less	Indirect business taxes	25
Less	Business transfer payments	1
Less	Current surplus of government enterprises	1
Plus	Subsidies	4
Equals	National income	$205

PERSONAL INCOME

As was outlined in Chapter 2, the format above, which shows the relationships among the three measures of output, leads directly to the derivation of another measure, *personal income*, which is defined as the current income of persons or households from all sources.[21] Unlike the other measures, personal income is not a measure of production, for it both excludes some items that are matched by productive activity and includes other items that are not. Thus, it includes both receipts for the productive services provided by persons and receipts, such as transfer payments, for which no productive services were provided by the recipients. Personal income is, therefore, derived from national income by subtracting from national income whatever parts of national income do not become receipts of persons and by adding to this remainder whatever receipts persons derive from sources not included in national income. The steps involved may be shown as follows:

National income		$205
Less	Contributions for social insurance	10
Less	Corporate profits tax	10
Less	Undistributed corporate profits	4
Plus	Business transfer payments	1
Plus	Government transfer payments	13
Plus	Net interest paid by government	4
Plus	Interest paid by consumers	1
Equals	Personal income	$200

DISPOSABLE PERSONAL INCOME

It will be recalled from Chapter 2 that another aggregate measure is derived from personal income by deducting from that total the amount of personal income that is siphoned off by government in personal taxes. The remainder is *disposable personal income*, since it equals the amount available to persons to dispose of as they choose. Whatever part of this they choose not to use for *personal outlays* remains as *per-*

[21] Income here includes more than actual cash receipts of persons; imputed or noncash receipts are also included. The role of imputations in the measures of production and in personal income is considered in Section V.

sonal saving.[22] In the three-sector economy, personal outlays equal personal consumption expenditures and interest paid by consumers. These relationships are shown below.[23]

Personal income			$200
Less	Personal taxes		24
Equals	Disposable personal income		$176
Less	Personal outlays		161
	Personal consumption expenditures	$160	
	Interest paid by consumers	1	
Equals	Personal saving		$ 15

GROSS NATIONAL PRODUCT—FROM SECTOR OF ORIGIN TO SECTOR OF EXPENDITURE

In measuring national income, net national product, and gross national product, we obtained each total by combining the national income or the net or gross national product originating in each of the three sectors. While such an arrangement answers the question of how much of the total national product was produced by each sector, it does not tell us how much of the product each sector secures for its own use. To learn this, we must turn to the expenditures of each sector on product. The aggregate of the expenditures of all sectors (including change in inventories as an expenditure) is necessarily of the same magnitude as the aggregate of product originating in all sectors, since there is an accounting identity between the total produced by all sectors and the total secured by all sectors. Table 8 shows how the presentation of gross national product by *sector of origin*, or sector of supply, may be rearranged to show the same total by type of expenditure, or what may be called *sector of expenditure*. The left-hand side of Table 8 corresponds with the product side of Account 4 of Table 7, although the order of entries has been changed to facilitate this comparison.

Here we see that, from the point of view of gross national product by sector of origin, government accounts for only $20. As a sector of expenditure it secures $50. The difference of $30 between what government produces directly and what it secures in total is the amount of goods and services produced by the business sector that is purchased by the government sector. The same reasoning applies to the $155 excess of household consumption ($160) over household production ($5). Since both the government and the household sectors secure

[22]In the table on p. 25 of Chapter 2, the items were arranged to show the subtraction from disposable personal income needed to reach personal consumption expenditures. Here the arrangement subtracts the items that add up to personal outlays to leave personal saving as the remainder.

[23]In the actual estimating procedure, just as in the above, personal saving is secured simply as a residual by subtracting estimated personal outlays from estimated disposable personal income. Because it is a residual and of a magnitude that is small relative to the two estimates from which it is obtained, it is subject to much larger relative errors than the other two estimates. In the illustration above, if the correct figure for disposable personal income was $178 and that for personal outlays was $159, or an error of approximately 1 percent in each case, the correct figure for personal saving would be $19 so that the $15 figure would involve approximately a 21-percent error.

TABLE 8
Gross national product by sector of origin and by sector of expenditure

		SECTOR OF ORIGIN	SECTOR OF EXPENDITURE		
$ 20		Gross national product originating with government	Government purchases of goods and services		$ 50
	$ 20	Wages, salaries, and supplements ⟶	Services provided by government employees	$ 20	
			Goods and services purchased from business	30	
225		Gross national product originating with business	Gross investment		40
	30	Sales to government	Goods purchased from business on capital account	35	
	35	Sales to business on capital account			
	5	Change in inventories ⟶	Change in inventories	5	
	155	Sales to consumers	Personal consumption expenditures		160
5		Gross national product originating with households	Goods and services purchased from business	155	
	5	Wages, salaries, and supplements ⟶	Services purchased from households	5	
$250		Gross national product	Gross national product		$250

more than they supply or produce, it follows that the business sector must supply more than it secures. This should not be taken to mean that households and government get something for nothing. Recall that gross national product originating in the business sector equals the sum of capital consumption allowances, indirect taxes plus minor adjustments, and national income originating in that sector. Most of the national income originating in the business sector becomes personal income of the household sector, and this enables households to purchase from business the bulk of its output. Similarly, a portion of the gross national product originating in business, mainly corporate profits taxes and indirect taxes, is a receipt of the government sector. These receipts, augmented by taxes on personal income, enable government to purchase a large part of business output. That portion of gross national product originating in the business sector that is not secured by either government or households remains in the business sector as unconsumed output, or gross investment.

FUNDAMENTAL IDENTITIES IN THE THREE-SECTOR ECONOMY

We have seen that in the absence of government, gross national product would be divided into consumption and gross investment. On the product side of the account, whatever is not consumed is gross investment, and, on the income side of the account, whatever is not consumed is gross saving. Gross saving is thus necessarily identical with gross investment.[24]

These relationships are modified by the introduction of government. On the product side,

[24]See Account 3 of Table 4.

gross national product is split three ways because there are now three sectors that demand final goods and services: households, business, and government. Similarly, on the income side of the account, there is a three-way split, for income is now devoted to taxes as well as to consumption and saving.

In adding government spending to gross national product, only that part of government expenditures that goes toward the purchase of goods and services is included.[25] Therefore, in calculating gross national product, we exclude from government expenditures all transfer payments, interest payments, and subsidies minus current surplus of government enterprises. Since for every dollar on the product side there must be a dollar on the income side, we must similarly exclude from the income side an amount of taxes equal to these other government expenditures. In Account 3 of Table 6, government purchases of goods and services total \$50, and this amount properly appears as part of gross national product. Other government expenditures total \$20 and are not part of gross national product. Therefore, on the income side of the account, it is necessary to deduct \$20 from gross tax receipts to yield net tax receipts. In effect, the \$20 total of government transfers and related expenditures are negative taxes, for these government expenditures of \$20 simply restore to the income side in redistributed form the very amount withdrawn from that side through taxes in the amount of \$20.

Thus, by taking taxes as net of these other government expenditures, we establish an identity between consumption (C), saving (S), and taxes (T) on the income side, and consumption (C), investment (I), and government purchases of goods and services (G) on the product side, or

$$C + S + T = C + I + G$$

This will be recognized as the identity that was given in Chapter 2 for the three-sector economy.

In any given time period, the amount of government purchases of goods and services on the product side, or G, may be equal to, greater than, or less than the amount of taxes, T, on the income side. In other words, the government may show, respectively, a balanced budget, a deficit, or a surplus. How each of these three possibilities affects the fundamental identity was sketched in Chapter 2; here a more detailed examination is provided.

Balanced Budget

Assume the budget is balanced with total government expenditures of \$69; purchases of goods and services are \$50, other expenditures are \$19, and total taxes are \$69. Limiting expenditures to those for goods and services and adjusting taxes correspondingly to a net basis gives us a budget balanced with G of \$50 and T of \$50. Thus, the amount that government adds to the stream of spending for the period through its purchases of goods and services is exactly matched by the amount that it withdraws from the stream of income generated by the stream of spending on final goods and services.

The fundamental identity including government is $C + S + T = C + I + G$. Since with a balanced budget $T = G$, it follows that $C + S = C + I$ and, from this in turn, that $S = I$. In other words, with a balanced budget, the identity between gross saving of the household and business sectors on the income side and gross investment of the business sector on the product side is just as it was in the two-sector economy. For every dollar of personal and business saving, S, there is a dollar of gross investment, I.

[25] It will be recalled that gross national product as a measure of the value of final product can be obtained by summing up all expenditures on final product, including in expenditures on final product any change in inventories. To include government expenditures for other than final product is to overstate gross national product by that amount and to destroy its meaning as a measure of the value of final output of goods and services.

Deficit

For this case we can make use of the figures in Table 6, in which government shows a deficit. Here, total government expenditures of $70 exceed total taxes of $69, producing a deficit of $1. Government purchases of goods and services, *G*, are $50, and taxes on a net basis, *T*, are $49. There thus remains on the income side $1, which was put there by *G* but was not withdrawn by *T*. This remaining $1 on the income side is accounted for within the time period as $1 of additional *S*, or saving by persons and firms. Thus, we have:

$$C + S + T = C + I + G$$
$$\$160 + \$41 + \$49 = \$160 + \$40 + \$50$$

and

$$S + T = I + G$$
$$\$41 + \$49 = \$40 + \$50$$

Note that, with an unbalanced budget, the sum of personal and gross business saving, *S*, no longer equals gross investment, *I*, for *S* is $41 and *I* is $40. Yet it has been repeatedly emphasized that gross saving must equal gross investment as an unavoidable accounting identity, since investment is unconsumed product, saving is unconsumed income, and income equals the value of product. We must show that an unbalanced budget does not and cannot upset this fundamental identity.

Again in terms of the present set of figures, the sum of personal consumption expenditures, *C* ($160), and "public consumption expenditures," *G* ($50), measures total consumption of output in the economy.[26] Gross investment, *I* ($40), is the unconsumed output of the economy. For every dollar of investment there must be a dollar of saving. Yet we have $41 of *S* and $40 of *I*. This extra dollar of *S* matches the government deficit of $1, and it is the deficit that produces this *apparent* inequality between saving and investment. Once the concept of government saving is introduced into our economy, we will see that the apparent inequality between saving and investment is just that—an *apparent* and not an actual inequality.

As in the household sector, where the difference between disposable personal income and personal outlays equals personal saving, in the government sector we find that the difference between net taxes and government purchases of goods and services equals *government saving*. If the household sector's personal outlays are greater than its disposable income, it "dissaves" to that extent, and personal saving for that period is negative. If the government spends more on goods and services than it collects in net taxes, it also dissaves to that extent, and government saving for the period is negative. Suppose, now, that we put together personal saving and gross business saving (composed of capital consumption allowances and undistributed profits) and call the total *gross private saving*, as distinct from government saving, which we will call *public saving*. Gross saving for the economy as a whole is then the sum of gross private saving (S) and public saving ($T - G$), and it is this sum of $S + (T - G)$ that must equal gross investment. And it does. The sum of gross private saving ($41) and public saving (−$1) is $40, and gross investment is also $40.

If we rearrange the fundamental identity as follows, we see that this is the case.

$$S + T = I + G$$
$$\$41 + \$49 = \$40 + \$50$$
$$S + (T - G) = I$$
$$\$41 + (\$49 - \$50) = \$40$$

A look back at Account 5 of Table 6, from which these figures were drawn, gives another view of the identity. For every dollar of *I* and *G* on the product side, there is a dollar of *either S* or *T* on the income side. $I + G$ total $90; $S + T$ total $90. If government withdraws less from the

[26] Government purchases of goods and services are all treated as if they were goods and services consumed, despite the fact that these purchases include long-lived goods such as buildings, dams, and ships. The same issue is faced under personal consumption expenditures where purchases of cars, appliances, and other durable goods are treated as consumption expenditures in the same way that gasoline to power cars and electricity to power appliances are treated.

income stream through net taxes than it adds through its purchases of goods and services, the difference must appear in that time period as private saving and public dissaving.

Whatever the government deficit for any time period, there is necessarily an excess of private saving over investment in that period equal to the deficit. It then follows that government can, at least potentially, finance its deficit in any time period by borrowing an amount equal to the deficit from private saving, which will be larger than investment by the amount of the deficit. Alternatively, the government can finance its deficit by "printing" new money in the amount of the deficit. In this case, it pays for the excess of its expenditures over its taxes without borrowing from the public. But the amount of the deficit is still matched by an equal excess of private saving over investment. The difference is simply that in the case in which government borrows from the public, the public's saving is held in government securities, and in the case in which government finances its deficit with newly created money, the public's saving is held in the form of money. How the deficit is financed makes a great difference in other respects, especially in terms of the possible inflationary impact of the deficit, but, for the fundamental identity with which we are here concerned, either method gives the same result: *Gross private saving exceeds gross investment by the amount of the deficit.*

Surplus

In the case of a surplus, taxes withdraw from the stream of income generated by total spending on final product an amount greater than that which government purchases of goods and services contribute to this stream of income.

Let us assume an economy with a surplus of \$1. Suppose government purchases of goods and services are \$50 as before and net taxes are now \$51. The identity reads:

$$C + S + T = C + I + G$$
$$\$160 + \$39 + \$51 = \$160 + \$40 + \$50$$

The government surplus, or public saving, of \$1 is necessarily offset by private saving, which is \$1 *less* than it would otherwise be. For the economy, gross saving of \$40 (\$39 private and \$1 public) is matched by gross investment of \$40. Again, by dropping C from both sides and by rearranging the identity, we have:

$$S + (T - G) = I$$
$$\$39 + (\$51 - \$50) = \$40$$

In this case, \$1 more is withdrawn from the income side by government than is in effect placed on that side by government purchases of goods and services. This extra \$1 is accounted for within the time period by a decrease of \$1 in private saving. The economy's gross investment of \$40 is matched dollar for dollar by gross saving. In the case of a surplus, part of the economy's saving is by government ($T - G$), so private saving (S) is necessarily less than investment by the amount of government saving.

To summarize, for all three possible cases—balanced budget, deficit, or surplus—the economy's three sectors must collectively, in any time period, show an amount of saving (or unconsumed income) equal to that period's investment (or unconsumed product). The three cases differ only in that each gives a different distribution of the economy's gross saving between private and public saving. A government deficit must add to private saving an amount equal to the deficit, for government is withdrawing from the income side less than it places there through its purchases of goods and services on the product side. A government surplus must subtract from private saving an amount equal to the surplus, for government is withdrawing more from the income side than it places there through its purchases of goods and services on the product side. A surplus or deficit, in other words, so alters gross private saving in the given time period that private saving, when added to government saving or dissaving, yields a total equal to the period's gross investment. With a balanced budget, gross private saving itself equals gross investment.

CIRCULAR FLOW IN THE THREE-SECTOR ECONOMY

The circular flow of income and product for the three-sector economy is given in Figure 2. The dollar values shown are those in Table 6. The gross national product of $250 is broken down by sector of origin to the left of center and by sector of expenditure to the right of center. On the left of the figure we have a flow of income of $250, which results from the economy's flow of product for the period.[27] To this flow of income created by production is added a flow of income to which there corresponds no production—namely, an income flow of $20 that results from government transfer payments and related items. Since these items are not part of the economy's production, they appear at the very outer edge of the figure, separated from the economy's gross national product. In this broader sense, the economy's income totals $270, as seen at the far left of the figure. This total of $270 of income is disposed of by its recipients in three ways: Part is spent for con-

[27]In the two-sector economy we assumed for simplicity that product originated only with the business sector. The circular flow, accordingly, was simply one of product from firms to households and of productive services from households to firms. In the present three-sector economy, we are allowing for the fact that production takes place in all three sectors, so a diagram like Figure 1 is no longer appropriate to illustrate the circular flow.

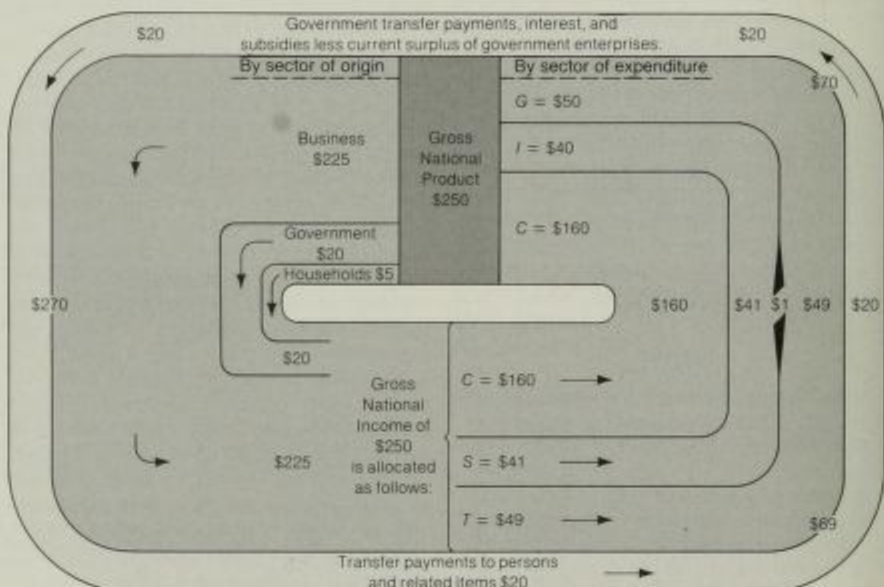


FIGURE 2
Circular flow in the three-sector economy

sumer goods, $160; part is taken by government in taxes, $69; and the remainder is saved, $41.

In order to trace the fundamental identity between the total of *C*, *S*, and *T* on the left and the total of *C*, *I*, and *G* on the right, it is again necessary to put taxes on a net basis by subtracting $20 from taxes of $69.[28] This subtraction reduces taxes to $49, and the sum of *C*, *S*, and *T* to $250 on the left. This is matched by the $250 sum of *C*, *I*, and *G*, or gross national product by sector of expenditure, on the right. Alternatively, we could arrive at the same result in Figure 2 by ignoring altogether government transfer payments and related government expenditures as a source of income. This would limit income on the left to that earned in the course of production, or $250, and it would limit spending on the right to expenditures on final goods and services, or $250. This $250 may thus be split up into *C*, *S*, and *T* on the left, which matches the sum of *C*, *I*, and *G* on the right without any adjustment.

The circular flow for this three-sector economy shows a government deficit of $1 for the time period. As discussed earlier, this means that government purchases of goods and services, *G* of $50, exceed net taxes, *T* of $49, by $1, and this $1 is shown as $1 of private saving in excess of gross investment. Thus, Figure 2 shows a flow of private saving of $41 and of investment of $40 and a flow of net taxes of $49 ($69 − $20) and of government purchases of goods and services of $50. On the right of the figure a small channel appears through which $1 is diverted from the stream of private saving into the stream of taxes. To show the necessary balance between $S + T$ and $I + G$, one might think of this as $1 of private saving that was borrowed by government to finance the excess of its purchases of goods and services over its net tax receipts.

III
NATIONAL INCOME ACCOUNTING: FOUR-SECTOR ECONOMY

An economy whose foreign transactions are excluded from analysis is termed a "closed" economy. By now including transactions with the "rest-of-the-world," we will have an "open" economy that includes all four sectors found in any actual economy.

The addition of the rest-of-the-world as the fourth and final sector now permits us to recognize that an actual economy's GNP is not necessarily equal to the total of final goods and services secured by domestic consumers (as measured by personal consumption expenditures), domestic government (as measured by government purchases of goods and services), and domestic business (as measured by gross private domestic investment). If an economy for any time period shows *net exports* of goods and services, the amount of final goods and services *secured* by domestic consumers, government and business is necessarily less than the amount domestically *produced*. If, however, the economy shows *net imports*, then domestic consumers, government, and business secure a total of final goods and services greater than that which was domestically produced during the time period. The former situation has been

[28]As discussed earlier, this subtraction is to offset the subtraction of $20 from total government expenditures of $70, the latter being necessary to show only those government expenditures of $50 to which there corresponds production and which are therefore properly part of gross national product.

much more common for the United States, although there have been years, including several recent ones, in which imports exceeded exports.[29]

As national income accounting seeks to provide estimates of the amount of final goods and services produced by the economy, it follows that the estimate for that total will increase from one period to the next with an increase in net exports or a decrease in net imports, *other things being equal*. It will decrease with a decrease in net exports or an increase in net imports, *other things being equal*. In this section we will examine the way that the net export or net import balance and some related items fit into the national income and product accounts.

NET FOREIGN INVESTMENT

Since an economy's income and product accounts measure, among other things, that economy's income and product, foreign transactions enter into its accounts only to the extent that they influence its income and product. The producing and consuming, saving and investing, and importing and exporting that take place in the rest-of-the-world have no effect on the accounts of the domestic economy, unless such activities involve transactions with the domestic economy. The introduction of the rest-of-the-world in no sense measures income or product in the rest-of-the-world; it shows only the effect on the home economy's income and product of transactions between the home economy and the rest-of-the-world.

It is true, of course, that whatever changes occur in the domestic economy's accounts through its foreign transactions must be matched by offsetting changes in the combined accounts of the rest-of-the-world, which record its transactions with the domestic economy. If we assume at this point that the only transactions between countries are purchases and sales of goods and services, it turns out that a single item enters to balance the international books: *net foreign investment* (or disinvestment) by the domestic economy in the rest-of-the-world is matched by *net foreign disinvestment* (or investment) by the rest-of-the-world in the domestic economy.

An economy's net sales of goods and services to the rest-of-the-world are called "net foreign investment"; they are goods and services produced, but not secured, by the domestic economy's nationals. In this sense, net foreign investment is similar in nature to domestic investment as developed earlier for a closed economy. Investment there was simply defined as the difference between what was produced and what was consumed, or as the amount of unconsumed output. Now, when we distinguish between domestic investment and foreign investment, we find that the difference lies solely in the disposition of the economy's unconsumed output. Domestic investment is that part of unconsumed output that remains at home in the form of an addition to inventories or to the stock of capital goods; foreign investment is that part of unconsumed output that goes abroad, where, depending in part on the nature of the goods, it may be consumed, added to inventories, or added to the stock of capital goods of the rest-of-the-world.[30]

[29] For the 48-year period from 1929 through 1976, measured in 1972 dollars, the U.S. economy showed net imports of goods and services in eleven years: 1935–37, 1942–45, 1968–69, and 1971–72.

[30] Net purchases from the rest-of-the-world are net foreign disinvestment to the domestic economy. If, however, the net purchases are not consumed but are added to inventories or to the stock of capital goods, domestic investment of the economy rises by the amount of its net foreign disinvestment, thus leaving the sum of domestic

How net foreign investment fits into the accounts may be seen in general terms by expanding the fundamental identities developed for the three-sector economy. For that economy, the identity for GNP was

$$C + S + T = C + I + G$$

Total product by sector of expenditure was composed of goods and services consumed by households (C) or by government (G), or left unconsumed (I). Now, with the addition of the fourth sector, some portion of total domestic production is sold to the rest-of-the-world and is available neither for consumption by households or government nor for domestic investment. At the same time, some portion of the total amount of goods and services consumed by households and by government at home and some portion of the total amount of capital goods purchased and goods added to business inventories at home were not produced domestically but were purchased from the rest-of-the-world. Gross national product is accordingly equal to $C + I + G$ plus the excess of sales to foreigners over purchases from foreigners. This excess is termed *net foreign investment* and is designated I_f. Designating gross domestic investment I_d, we have the following identity for GNP in the four-sector economy:

$$C + S + T = C + I_d + G + I_f$$

Since net foreign investment, I_f, equals sales abroad less purchases abroad, we may also express the identity as follows:

$$C + S + T = C + I_d + G + (SA - PA)$$

in which SA designates sales abroad and PA purchases abroad.

Both versions of this fundamental identity differ from the fundamental identity for the three-sector economy only in the breakdown on the product side. Unconsumed output in the three-sector economy was limited to gross domestic investment, whereas unconsumed output in the four-sector economy is the sum of gross domestic and net foreign investment.

For the three-sector economy, we derived the identity between saving and investment from $C + S + T = C + I + G$ by canceling the C's and rewriting to yield $S + (T - G) = I$. Here the economy's gross saving (private saving of S and public saving of $T - G$) equals the economy's gross investment. Now, for the four-sector economy, gross saving remains the sum of private and public saving, but gross investment becomes the sum of gross domestic investment, I_d, and net foreign investment, I_f. The saving–investment identity now reads

$$S + (T - G) = I_d + I_f$$

or

$$S + (T - G) = I_d + (SA - PA)$$

since $(SA - PA)$ equals I_f.

FOREIGN TRANSFER PAYMENTS, CAPITAL GRANTS, AND INTEREST PAID BY GOVERNMENT TO FOREIGNERS

The gross national product identity just developed for the four-sector economy includes under $C + I_d + G + (SA - PA)$ all the final goods and services produced during the period, but it does not accurately indicate the actual portion of final goods and services that was secured by the rest-of-the-world. This apparent contradiction results from the fact that the domestic economy's net foreign investment measures only the *net purchases* by the rest-of-the-world from the domestic economy. However, in addition to purchases, for which by definition payment is financed by the pur-

and net foreign investment unchanged. Similarly, net sales to the rest-of-the-world are net foreign investment. If, however, the amount of net sales represents a net withdrawal from the economy's inventories or stock of capital goods, domestic disinvestment offsets net foreign investment, thus leaving the sum of domestic and net foreign investment unchanged.

chaser, goods and services are exported from the domestic economy for which payment is not so financed and which are thus not included in purchases. Yet the net amount of goods flowing from the domestic economy is part of that economy's GNP, even though this net amount does not appear in full as net purchases by the rest-of-the-world. This difference must be included as part of GNP and is so included under the headings of government purchases of goods and services and personal consumption expenditures.

For example, suppose that the federal government in a given year makes $2 billion in *grants* to other countries in the form of goods. This $2 billion would not appear as a part of net purchases from the United States and therefore would not be included in net foreign investment, but it would appear as a part of government purchases of goods and services, *G*. Similarly, suppose that in a given year persons give $1 billion in the form of goods to the rest-of-the-world. This, again, does not appear as part of net purchases from the United States and therefore is not included in net foreign investment; rather, it appears as part of personal consumption expenditures, *C*.

If our purpose is only to measure the economy's income and product for the year, this procedure is altogether appropriate, for it does pick up all final goods and services. If we wish also, however, to measure the division of the GNP into goods and services secured by the nationals of the domestic economy and goods and services secured by the rest-of-the-world, this procedure is not altogether appropriate.

Before World War II, this approach was not seriously inappropriate for the latter purpose, since in those years government grants were insignificant in comparison with what they have been since the war. In view of the postwar rise in government grants, the Commerce Department in 1958 introduced a revised treatment of foreign transactions to remove from government purchases and classify as foreign transfer payments an amount equal to net cash grants to the rest-of-the-world and thus to show this amount explicitly as part of foreign transactions. In the revisions introduced in 1965, government nonmilitary grants and personal remittances in cash and in kind have been accorded the same treatment.[31] In the household-sector account, a new entry—personal transfer payments to foreigners—appears, the amount of which is no longer included in personal consumption expenditures. Since, as a result of these changes, foreign transactions are no longer limited to *sales* and *purchases*, it is appropriate to use the broader terms *exports* and *imports*. Although net exports (exports minus imports) now exceed net foreign investment, this causes no change in the magnitude of GNP. Apart from a minor adjustment for capital grants to be considered in the following paragraph, the amount by which net exports ($X - M$) exceed net foreign investment ($SA - PA$) is offset by a reduction in what otherwise would have been larger figures for *G* and *C*. The amount by which *G* is reduced is an amount called *transfer payments from U.S. government to foreigners (net)*, and the amount by which *C* is reduced is an amount called *personal transfer payments to foreigners (net)*.

Another revised treatment became necessary in 1970 with the activation of the Special Drawing Rights (SDR) system of the International Monetary Fund (IMF). In January 1970, $3,414 million of such rights were entered to the credit of the member countries in the Special Drawing Account of the IMF. The quota for each country was proportional to its IMF quota, and the U.S. share of this initial issue was approximately $0.9 billion. Its share of the 1971 and 1972 issues was approximately $0.7 billion in each year. No additional SDRs have been issued since 1972. In 1974 the U.S. used $2.0 billion

[31] Military grants continue to be classified as government purchases of goods and services. Thus, it is the purchase of military equipment by the U.S. government that appears in the accounts rather than its subsequent transfer to foreign nations.

or almost all of its accumulated holdings of SDRs in settlement of international debt. The rights themselves are something created out of thin air and involve no more than bookkeeping entries in the IMF Special Drawing Account. They exist on paper only and hence are popularly called "paper gold." Unlike gold, their value as an international reserve asset to each country rests solely on the agreement among all countries to accept them in settlement of international balances.

Because each member country gets its allocation without putting up anything in exchange, SDRs represent additions to the foreign assets of a country that are not matched by a corresponding addition to its liabilities. It therefore appears appropriate to count these as part of a nation's net foreign investment during the period in which they are received. In the case of the United States, the effect of the initial issue of SDRs was, therefore, to increase by $0.9 billion whatever the net investment figure for 1970 would otherwise have been. It is then necessary to show a corresponding amount in the accounts as the source of this increase in net foreign investment. Because the allocation of SDRs is in the nature of a unilateral transfer or grant to the government from an international agency, it could be netted out in the U.S. accounts by an equal reduction in the amount under transfer payments from U.S. government to foreigners. However, that item in practice has included only current transfers whereas SDRs represent a capital transfer. The Commerce Department resolved the problem of accounting for the source of SDRs by introducing a new item called *capital grants received by the United States*. This item so far includes nothing other than the allocation of SDRs. For years in which grants were received from the IMF, the item is positive; for years in which such amounts were used to make payments, the item is negative.

Yet another revised treatment was introduced in 1976. Up till then interest paid by government to foreigners had been classified as government purchases of goods and services and as imports of goods and services.[32] As the inclusion of the amount in question under government purchases, a positive item, was offset by its inclusion under imports, a negative item, the amount was not included in GNP. Because of the great increase in recent years in the amount of federal government securities held by foreign investors, OPEC countries and others, government interest payments to these holders have run into billions of dollars. Due to the amounts involved, the Department of Commerce judged it inadvisable to continue the old treatment. The new treatment removes this amount from government purchases and shows it as part of net interest paid on the expenditures side of the government-sector account. (See Account 3 of Table 6.) This amount is also deducted from imports, thereby increasing net exports by the same amount. The increase in net exports offsets the decrease in government purchases to leave GNP unchanged by this revised treatment. As will be seen below, the subtraction of this amount from imports is balanced in the account for foreign transactions by showing that same amount under the new item: interest paid by government to foreigners.

FOREIGN TRANSACTIONS ACCOUNT

This treatment of the various items yields what is officially called the Foreign Transactions Account, shown as Table 9. With foreign transfer payments and capital grants as there given, the

[32]This treatment constituted an exception to the rule noted in footnote 12. Although interest paid by government was, in general, excluded from government purchases, this exception was made to secure uniformity with the treatment of this flow in the balance of payments accounts.

$5 excess of exports over imports is covered by $2 of foreign transfer payments from government, $1 of foreign transfer payments from persons, and $2 of net foreign investment. However, in addition to the receipts from exports, there is the additional receipt of $1 from capital grants. Exports plus capital grants exceed imports by $6, and the $1 of capital grants makes net foreign investment the amount of $3 as shown instead of the $2 it would be if capital grants were zero.

As for exports and imports of goods and services, they here have the usual meanings. Included are familiar items such as commodities of all kinds and "invisibles" such as shipping, banking, insurance services, and tourism, although this detail is not shown in the highly condensed Foreign Transactions Account. For national income purposes, the specific services described as factor services are separated out to determine the amount of national income originating in this sector. Income earned for such services includes wages and salaries paid to nationals of the domestic economy who are employed either at home or abroad by foreign firms, governments, and institutions; interest and dividends received by nationals on holdings of securities issued by foreign companies and governments; and profits earned by foreign branches of domestic firms. The domestic economy also purchases such services from the rest-of-the-world and accordingly makes similar payments to the rest-of-the-world. If a breakdown of Table 9 showed $5 for factor services under exports of goods and services and $1 for factor services under imports of goods and services, on a net basis, sale of factor services would amount to $4, and this would be the amount of the domestic economy's national income originating with the rest-of-the-world. National income, it will be recalled, is a measure of the earnings of the domestic economy's factors of production, and the net amount earned by these factors through providing such services to the rest-of-the-world is measured in this way and combined with the amounts earned by the factors from the other three sectors to yield the measure of the total national income for the economy.

It would now be possible to develop an entire set of accounts for the four-sector economy with all the detail given in Table 6 for the three-sector economy. However, we can bypass this step and still see how each of the entries in the Foreign Transactions Account fits into the full set of accounts by simply indicating the counterentry for each as it would appear in the complete set of accounts.[33] Net exports of $5 in Table 9 has as its counterentry the same item on the product side of the National Income and Product Account. Net foreign investment has as its counterentry the same item under gross investment in the Gross Saving and In-

[33]These same entries may be traced in Table 11 (in the following section), which presents the full set of summary accounts with data for 1976 for the U.S. economy.

TABLE 9
Foreign transactions account

Exports of goods and services	$23	Imports of goods and services	$18
Capital grants received	1	Transfer payments from domestic government to foreigners (net)	1
		Personal transfer payments to foreigners (net)	1
		Interest paid by government to foreigners	1
		Net foreign investment	3
Receipts from foreigners	$24	Payments to foreigners	$24

vestment Account. Foreign transfer payments of government and of persons each has as its counterentry the same item on the allocations side of the government-sector and household-sector accounts. Interest paid by government to foreigners has as its counterentry the same item on the allocations side of the government-sector account. Lastly, capital grants received by the United States has as its counterentry the same item on the saving side of the Gross Saving and Investment Account. Because this last amount is something received by the United States without payment of any kind by government, business, or persons within the United States, it is best viewed as an equal amount of saving. It is, however, not counted as part of private (personal plus business) or public saving, because the way it is handled does not involve it explicitly in the accounts of any of these sectors. It is a type of saving that must be added to private and public saving to get the total for gross saving.

The fundamental identities developed earlier in this chapter must now be modified to allow for the introduction of foreign transfer payments and capital grants. We had earlier the following for GNP:

$$C + S + T = C + I_d + G + I_f$$

or

$$C + S + I = C + I_d + G + (SA - PA)$$

Now instead of showing $SA - PA$ or *net sales* as the amount of spending accounted for by foreign transactions, we show $X - M$, which exceeds $SA - PA$ in that it allows for the fact that goods may go abroad that are not "sold" but are rather financed by the domestic economy's transfer payments to foreigners.

If we designate transfer payments from domestic government to foreigners by R_{gf}, transfer payments from persons to foreigners by R_{pf}, government interest paid to foreigners by R_{gif}, and capital grants received by the domestic economy by R_{cg}, we may express the identity in the Foreign Transactions Account as shown in the next column.

$$X + R_{cg} = M + I_f + R_{gf} + R_{pf} + R_{gif}$$

For net exports, we then have

$$X - M = I_f + R_{gf} + R_{pf} + R_{gif} - R_{cg}$$

Reflecting the definitions underlying the items in the Foreign Transactions Account, GNP is now defined as the sum of $C + I_d + G + (X - M)$. Substituting $I_f + R_{gf} + R_{pf} + R_{gif} - R_{cg}$ for $X - M$, we may write the GNP identity as follows:

$$C + R_{pf} + S + T$$
$$= C + I_d + G + I_f + R_{gf} + R_{pf} + R_{gif} - R_{cg}$$

We earlier had the following identity for saving and investment:

$$S + (T - G) = I_d + I_f$$

or

$$S + (T - G) = I_d + (SA - PA)$$

Now, however, including foreign transfer payments by government, foreign transfer payments by persons, government interest paid to foreigners, and capital grants to the domestic economy, we derive the following identity for saving and investment from the GNP identity:

$$S + (T - G - R_{gf} - R_{gif}) + R_{cg} = I_d + I_f$$

in which S as before is gross private saving, $(T - G - R_{gf} - R_{gif})$ is public saving, and R_{cg} is a separate component of gross saving not included with either private or public saving.[34]

The full series of fundamental identities developed through the successive economies may

[34]Note that in the GNP identity above, R_{cg} appears as a negative item on the product side of the identity. The meaning of this can perhaps now be seen more clearly in this way. The treatment of this item initially involves showing R_{cg} on the income side as a part of gross saving (but not included as part of private or public saving) and on the product side as an amount of product in the form of additional net foreign investment or I_f. However, in this form both sides overstate the correct figure for GNP by the amount of R_{cg}. Foreign transactions are reflected in GNP by an amount equal to $X - M$, and $I_d + R_{gf} + R_{pf} + R_{gif}$ now exceeds $X - M$ by the amount of R_{cg}. Bringing R_{cg} over to the product side as a minus item makes $I_d + R_{gf} + R_{pf} + R_{gif} - R_{cg}$ equal to $X - M$ as required for the GNP identity.

TABLE 10

ECONOMY	GROSS NATIONAL PRODUCT IDENTITIES	SAVING AND INVESTMENT IDENTITIES
Two-sector	$C + S = C + I$	$S = I$
Three-sector	$C + S + T = C + I + G$	$S + (T - G) = I$
Four-sector	$C + R_{pf} + S + T = C + I_d + G + (X - M)$	$S + (T - G - R_{gf} - R_{gif}) + R_{cg} = I_d + I_f$

now be placed in order for comparison, as shown in Table 10.

In the two-sector economy, gross private saving was necessarily equal to gross domestic investment; there was neither public saving nor net foreign investment. In the three-sector economy, the sum of gross private saving and public saving was necessarily equal to gross domestic investment; there was no net foreign investment. Finally, in the four-sector economy, the sum of gross private saving (S), public saving ($T - G - R_{gf} - R_{gif}$), and R_{cg} necessarily equals the sum of gross domestic investment and net foreign investment. Our final hypothetical economy has carried us to the point at which each class of saving and each class of investment found in the real world appears in the fundamental identities. However, from the two-sector to the four-sector economy, the principle is unchanged. In terms of GNP, for every dollar of final product on the right there is a dollar of income on the left; in terms of saving and investment, for every dollar of final product on the right that is not in the form of personal or government purchases (i.e., for every dollar of investment), there is a dollar of income on the left not spent for public or private consumption (i.e., a dollar of saving).

CIRCULAR FLOW IN THE FOUR-SECTOR ECONOMY

The circular flow for an open economy, shown in Figure 3, contains only a few modifications of Figure 2, the flow for a closed economy. In the upper part of Figure 3, under gross national product *by sector of expenditure,* the rest-of-the-world appears as a fourth sector, with its demand measured by net exports ($5); under gross national product *by sector of origin,* the rest-of-the-world appears as a fourth sector, with income originating abroad measured by net factor income received by the domestic economy from the rest-of-the-world ($4).

The dollar figures in Figure 3 correspond with the following hypothetical figures for the basic identities in the four-sector economy.

The gross national product identity is:

$$C + R_{pf} + S + T = C + I_d + G + (X - M)$$

or

$$C + R_{pf} + S + T$$
$$= C + I_d + G + I_f + R_{gf} + R_{pf} + R_{gif} - R_{cg}$$
$$\$159 + \$1 + \$41 + \$49$$
$$= \$159 + \$36 + \$50 + \$3 + \$1 + \$1 + \$1 - \$1$$

The saving and investment identity is:

$$S + (T - G - R_{gf} - R_{gif}) + R_{cg} = I_d + I_f$$
$$\$41 + (\$49 - \$50 - \$1 - \$1) + \$1 = \$36 + \$3$$

In Figure 3, GNP by sector of origin shows

a flow of \$250 of final goods and services, of which \$221 is assumed to originate with business, \$20 with government, \$5 with households, and \$4 with the rest-of-the-world. Corresponding to this flow of gross product is a flow of gross income, of which \$159 is devoted to personal consumption expenditures, \$41 to private saving, and \$1 to personal transfer payments to foreigners and of which \$49 is net taxes paid to government. In terms of spending for final goods and services, \$159 of these goods and services is secured by households, \$50 by government, \$36 by business in the form of capital goods and additions to inventories, and \$5 by the rest-of-the-world in the form of net exports from the domestic economy.

The essential difference between this circular flow and the flow for the three-sector economy lies in this breakdown of expenditures for final product. In the new flow we have the fourth sector of expenditure, net exports of \$5, here shown to be equal to net foreign investment of \$3 plus the sum of foreign transfer payments, and government interest paid to foreigners, and capital grants, which is \$2. This division permits the further division within the diagram of the economy's gross investment of \$39 into gross domestic investment of \$36 and net foreign investment of \$3. From the disposition of gross income, we find that gross private saving is \$41, or \$2 in excess of the sum of gross domestic and net foreign investment. At the same time, the sum of government purchases of goods and services, government transfer payments to foreigners, and government interest paid to foreigners is \$52, or \$3 greater than net tax re-

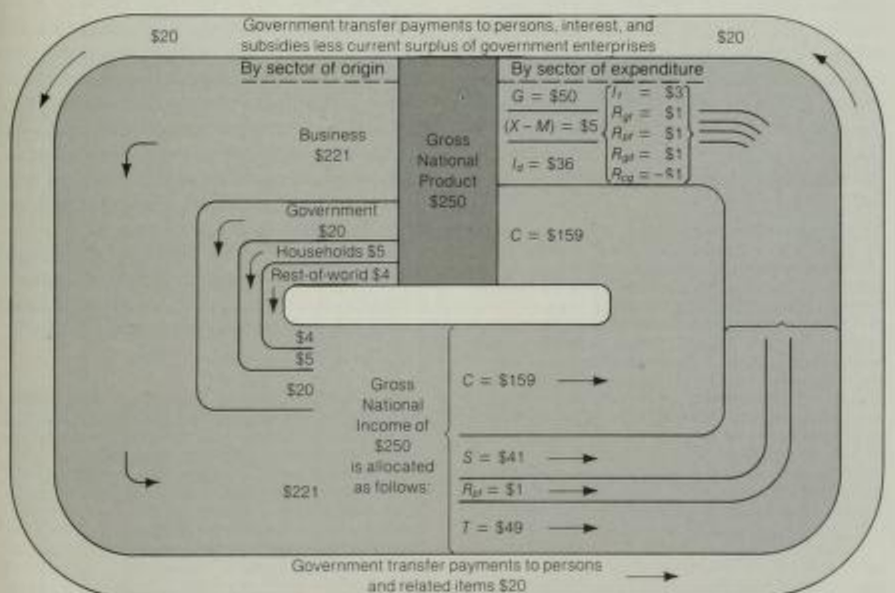


FIGURE 3
Circular flow in the four-sector economy

ceipts, so that government or public saving is −$3. This makes the sum of private and public saving $38. Adding the $1 of saving that is in the form of capital grants to this $38 brings gross saving up to $39 or equal to gross investment.

Apart from its hypothetical dollar amounts and the omission of some detail, Figure 3 gives one version of the circular flow of income and product for a real-world, four-sector economy. The step-by-step buildup of the circular flow process carried out through three successive hypothetical economies has brought us in Figure 3 to the border of the real-world economy of the United States. In the next section, we will examine a more detailed circular flow diagram that describes the U.S. economy in terms of the complete official accounting framework, incorporating the actual figures recorded for the year 1976.

IV THE NATIONAL INCOME AND PRODUCT ACCOUNTS OF THE UNITED STATES

Having worked through a series of increasingly complex accounting frameworks in the preceding sections, we turn here to the complete official accounting framework for the U.S. economy as constructed by the Department of Commerce. This is given in Table 11 with data for the year 1976.[35] The National Income and Product Account, the first account in the table, was reviewed in Chapter 2, but the other accounts are those for which the previous sections of this appendix have laid the groundwork.[36]

To understand fully the positioning of individual items and the overall organization of these items within the official five-account system, it is necessary to trace each item in the National Income and Product Account to its counterentry in the other accounts and to fit each item and its counterentry into a picture of the circular-flow process. For example, take wage and salary disbursements of $891.8 billion, the first item in the National Income and Product Account for which there is a counterentry. The counterentry, 2-7, is wages and salaries of this amount received by the household

[35]This system of accounts departs in several ways from the one built in the preceding sections. The official system does not include a separate business-sector account. This omission involves a sacrifice of some detail, but since the national income and product account is not too different from the business-sector account (compare, for example, Accounts 1 and 4 of Table 6), the loss of detail is not great and is compensated for by the simplification of the system. The official system of accounts also does not show national income and gross national product by sector of origin. From the data given in Table 11, it is not possible to break down national income or gross national product into the amounts originating in each of the four sectors of the economy. Although this detail is given elsewhere in the department's statistical tables, from the summary shown in Table 11 all we can do is break down national income into types of income and gross national product into types of expenditure. Like the omission of the entire business-sector account, the omission of items showing the institutional origins of national income and gross national product on the income side of the sector accounts simplifies the accounts. It also focuses attention on the analytically more useful breakdown of gross national product by sector of expenditure, as found on the product side of the National Income and Product Account.

[36]Note that certain items in Table 11 are designated by symbols; these will be used in the construction of identities like those in Chapter 2 but here including some detail there omitted. This system of symbols combines similar items under the same capital letter and uses subscripts to distinguish separate items in each class. Thus, repeating some earlier identified items, transfer payments and other payments of a transfer nature are all designated by R. Business transfer payments are identified by the subscript $_b$, government transfer payments to persons by $_{gp}$, government transfer payments to foreigners by $_{gf}$, government interest payments to domestic recipients by $_{gd}$, government interest paid to foreigners by $_{gf}$, personal interest payments by $_{pf}$, subsidies by $_s$, and capital grants or transfers by $_{cg}$. Similarly for various types of taxes, T_i, T_c, T_p, and T_s; saving, S_p, S_u, S_g, and S_g; and investment, I_g and I_f.

TABLE 11
Summary national income and product accounts, 1976[a]
(billions of dollars)

1.—National income and product account

ITEM			
1		Compensation of employees	1,036.3
2		Wages and salaries	891.8
3		Disbursements (2–7)	891.8
4		Wage accruals less disbursements (3–11) + (5–4)	0.0
5		Supplements to wages and salaries	144.5
6		Employer contributions for social insurance (3–19)	68.6
7		Other labor income (2–8)	75.9
8		Proprietors' income with inventory valuation and capital consumption adjustments (2–9)	88.0
9		Rental income of persons with capital consumption adjustment (2–10)	23.3
10		Corporate profits with inventory valuation and capital consumption adjustments	128.1
11		Profits before tax	156.9
12	(T_c)	Profits tax liability (3–16)	64.7
13		Profits after tax	92.1
14		Dividends (2–11)	35.8
15	(S_u)	Undistributed profits (5–6)	56.4
16		Inventory valuation adjustment (5–7)	−14.1
17		Capital consumption adjustment (5–8)	−14.7
18		Net interest (2–13)	88.4
19	(NI)	NATIONAL INCOME	1,364.1
20	(R_b)	Business transfer payments (2–19)	8.1
21	(T_i)	Indirect business tax and nontax liability (3–17)	150.5
22	(R_s)	Less: Subsidies less current surplus of government enterprises (3–10)	0.8
23	(SD)	Statistical discrepancy (5–12)	5.5
24		CHARGES AGAINST NET NATIONAL PRODUCT	1,527.4
25	(S_g)	Capital consumption allowances with capital consumption adjustment (5–9)	179.0
		CHARGES AGAINST GROSS NATIONAL PRODUCT	1,706.5

ITEM			
26	(C)	Personal consumption expenditures (2–3)	1,094.0
27		Durable goods	158.9
28		Nondurable goods	442.7
29		Services	492.3
30	(I_g)	Gross private domestic investment (5–1)	243.3
31		Fixed investment	230.0
32		Nonresidential	161.9
33		Structures	55.8
34		Producers' durable equipment	106.1
35		Residential	68.0
36		Change in business inventories	13.3
37	($X - M$)	Net exports of goods and services	7.8
38		Exports (4–1)	162.9
39		Imports (4–3)	155.1
40	(G)	Government purchases of goods and services (3–1)	361.4
41		Federal	130.1
42		National defense	86.8
43		Nondefense	43.3
44		State and local	231.2
		GROSS NATIONAL PRODUCT	1,706.5

[a]Numbers in parenthesis indicate accounts and items of counterentry in the accounts.

Summary national income and product accounts, 1976[a]—continued
(billions of dollars)

2.—Personal income and outlay account

ITEM				ITEM			
1	(T_p)	Personal tax and nontax payments (3–15)	196.9	7		Wage and salary disbursements (1–3)	891.8
2	(PO)	Personal outlays	1,119.9	8		Other labor income (1–7)	75.9
3	(C)	Personal consumption expenditures (1–26)	1,094.0	9		Proprietors' income with inventory valuation and capital consumption adjustments (1–8)	88.0
4	(R_{pb})	Interest paid by consumers to business (2–17)	25.0				
5	(R_{pf})	Personal transfer payments to foreigners (net) (4–5)	0.9	10		Rental income of persons with capital consumption adjustment (1–9)	23.3
6	(S_p)	Personal saving (5–3)	65.9	11		Dividends (1–14)	35.8
				12		Personal interest income	130.3
				13		Net interest (1–18)	88.4
				14		Interest paid by government to persons and business (3–7)	39.3
				15		Less: Interest received by government (3–9)	22.4
				16	(R_{gp})	Net interest paid by government to persons and business (3–7) – (3–9)	16.9
				17	(R_{pb})	Interest paid by consumers to business (2–4)	25.0
				18		Transfer payments to persons	192.8
				19	(R_{bp})	From business (1–20)	8.1
				20	(R_{gp})	From government (3–3)	184.7
				21		Less: Personal contributions for social insurance (3–20)	55.2
		PERSONAL TAXES, OUTLAYS, AND SAVING	1,382.7			PERSONAL INCOME	1,382.7

3.—Government receipts and expenditures account

1	(G)	Purchases of goods and services (1–40)	361.4	15	(T_p)	Personal tax and nontax payments (2–1)	196.9
2		Transfer payments	188.0	16	(T_c)	Corporate profits tax liability (1–12)	64.7
3	(R_{gp})	To persons (2–20)	184.7	17	(T_i)	Indirect business tax and nontax liability (1–21)	150.5
4	(R_{gf})	To foreigners (net) (4–6)	3.2	18	(T_s)	Contributions for social insurance	123.8
5		Net interest paid	21.4	19		Employer (1–6)	68.6
6		Interest paid	43.8	20		Personal (2–21)	55.2
7		To persons and business (2–14)	39.3				
8		To foreigners (4–7)	4.5				
9		Less: Interest received by government (2–15)	22.4				

11		Less: Wage accruals less disbursements (1–4)	0.0				
12	(S_g)	Surplus or deficit (–), national income and product accounts (5–10)	–35.6				
13		Federal	–54.0				
14		State and local	18.4				
		GOVERNMENT EXPENDITURES AND SURPLUS	535.9			GOVERNMENT RECEIPTS	535.9

4.—Foreign transactions account

1		Exports of goods and services (1–38)	162.9	3		Imports of goods and services (1–39)	155.1
2	(R_{cg})	Capital grants received by the United States (net) (5–11)	0.0	4		Transfer payments to foreigners (net)	4.2
				5	(R_{pf})	From persons (net) (2–5)	0.9
				6	(R_{gf})	From government (net) (3–4)	3.2
				7	(R_{gi})	Interest paid by government to foreigners (3–8)	4.5
				8	(I_f)	Net foreign investment (5–2)	–0.9
		RECEIPTS FROM FOREIGNERS	162.9			PAYMENTS TO FOREIGNERS	162.9

5.—Gross saving and investment account

1	(I_d)	Gross private domestic investment (1–30)	243.3	3	(S_p)	Personal saving (2–6)	65.9
2	(I_f)	Net foreign investment (4–8)	–0.9	4		Wage accruals less disbursements (1–4)	0.0
				5	(S_u)	Undistributed corporate profits with inventory valuation and capital consumption adjustments	27.6
				6		Undistributed corporate profits (1–15)	56.4
				7		Inventory valuation adjustment (1–16)	–14.1
				8		Capital consumption adjustment (1–17)	–14.7
				9	(S_d)	Capital consumption allowances with capital consumption adjustment (1–25)	179.0
				10	(S_g)	Government surplus or deficit (–), national income and product accounts (3–12)	–35.6
				11	(R_{cg})	Capital grants received by the United States (net) (4–2)	0.0
				12	(SD)	Statistical discrepancy (1–23)	5.5
		GROSS INVESTMENT	242.5			GROSS SAVING AND STATISTICAL DISCREPANCY	242.5

*Numbers in parenthesis indicate accounts and items of counterentry in the accounts.

SOURCE: *Survey of Current Business*, U.S. Department of Commerce, July 1977. The symbols have been added by the author.

sector as part of its personal income.[37] We may say that personal income is the amount it is *in part* because wages and salaries are the amount they are. Since personal consumption expenditures depend *in part* on the amount of income persons receive, and since that amount depends *in part* on wages and salaries received, we may say that personal consumption expenditures 2-3 of $1,094.0 billion are the amount they are *in part* because wage and salary disbursements 2-7 are the amount they are. Similarly, personal consumption expenditures 2-3 may in turn be traced to the counterentry, 1-26. At this point we may say that wage and salary disbursements 1-3, the starting point of our analysis, are *in part* the amount they are because personal consumption expenditures 1-26 are the amount they are. Alternatively we may backtrack to 2-7 and follow a more complicated route, saying that personal tax payments 2-1 of $196.9 billion are *in part* the amount they are because wage and salary disbursements 2-7 are the amount they are. The entry for personal tax payments 2-1 may in turn be traced to its counterentry, 3-15, and, assuming that government purchases 3-1 are not independent of tax receipts, we may say that government purchases of $361.4 billion are *in part* the amount they are because personal tax receipts 3-15 are the amount they are. The entry for government purchases may in turn be traced to its counterentry, 1-40, and, at this point, we may say that wage and salary disbursements 1-3, our original starting point, are *in part* the amount they are because government purchases 1-40 are the amount they are. Another route would be to backtrack to 2-7 and argue that personal saving 2-6 is the amount it is *in part* because wage and salary disbursements 2-7 are the amount they are, and then go on from there. In each case, we come full circle, tracing wage and salary disbursements 1-3 all through the system, right back to the point at which we started, 1-3. This may also be done for the other items in Account 1. Tracing these flows item by item is one way of tying together the circular relationships among the individual items out of which the accounts are built.

RELATION OF GROSS NATIONAL PRODUCT TO OTHER AGGREGATES

Tracing individual items through the accounts also reveals how each item fits into the various aggregates we have discussed in the preceding sections and in Chapter 2.[38] In our illustration, wage and salary disbursements 1-3, which are part of the aggregate national income, led to the counterentry of wage and salary disbursements 2-7, which is part of the aggregate personal income. Similarly, that part of national income that consists of undistributed profits, 1-15, leads to counterentry, 5-6, which is part of the aggregate gross saving. That part of GNP that consists of indirect business taxes, 1-21, leads to counterentry, 3-17, which is part of the aggregate government receipts.

[37] In some years wage and salary disbursements are less than wages and salaries as wage accruals less disbursements are positive. They were zero in 1976. Unlike the counterentry for disbursements, those for accruals less disbursements are 3-11 and 5-4. A part of the positive amount in a year may be accounted for by the business sector. This amount resembles undistributed corporate profits and may be positive or negative. Like undistributed profits, it appears as a form of business saving on the saving side of the Gross Saving and Investment Account. The other part of this amount, may be accounted for by the government sector, and it appears on the expenditure side of the Government Account.

[38] The term "aggregate" is used here to mean a combination of items that appear as separate items in Table 11. In a different sense, each of these items is itself an aggregate. Indirect taxes are an aggregate of state, local, and federal indirect taxes. Each of these in turn is an aggregate of sales taxes, property taxes, motor vehicle taxes, etc.

TABLE 12

Relation of gross national product, net national product, national income, personal income, and disposable personal income, selected years, 1929–1976 (billions of dollars)

		1929	1933	1940	1945	1950	1955	1960	1965	1970	1975	1976
The sum of												
	Personal consumption expenditures (C)	77.3	45.8	71.0	119.5	192.0	253.7	324.9	430.2	618.8	980.4	1,094.0
	Gross private domestic investment (I_g)	16.2	1.4	13.1	10.6	53.8	68.4	76.4	112.0	140.8	189.1	243.3
	Government purchases of goods and services (G)	8.8	8.2	14.2	82.8	38.5	75.0	100.3	138.3	218.9	338.9	361.4
	Net exports of goods and services ($X - M$)	1.1	0.4	1.7	−0.6	1.9	2.2	4.4	7.6	3.9	20.4	7.8
Equals	GROSS NATIONAL PRODUCT (GNP)	103.4	55.8	100.0	212.3	286.2	399.3	506.0	688.1	982.4	1,528.8	1,706.5
Less	Capital consumption allowances with capital consumption adjustment (S_g)	9.7	7.5	9.0	12.3	23.9	35.3	47.7	57.5	90.8	162.5	179.0
Equals	NET NATIONAL PRODUCT (NNP)	93.7	48.3	91.0	200.0	262.3	364.0	458.3	630.6	891.6	1,366.3	1,527.4
Less	Indirect business taxes (T_i)	7.1	7.1	10.1	15.5	23.4	32.2	45.4	62.6	94.0	138.7	150.5
	Business transfer payments (R_b)	0.6	0.7	0.4	0.5	0.8	1.2	2.0	2.8	4.0	7.0	8.1
	Statistical discrepancy (SD)	1.1	0.7	1.1	4.1	2.0	2.5	−0.7	0.9	−2.1	5.9	5.5
Plus	Subsidies minus current surplus of government enterprises (R_s)	−0.2	0.0	0.4	0.7	0.1	0.0	0.4	1.6	2.7	2.3	0.8
Equals	NATIONAL INCOME (NI)	84.8	39.9	79.7	180.6	236.2	328.0	412.0	566.0	798.4	1,217.0	1,364.1
Less	Corporate profits taxes (T_c)	1.4	0.5	2.8	10.7	17.9	22.0	22.7	30.9	34.5	50.2	64.7
	Contributions for social insurance (T_s)	0.2	0.3	2.3	6.1	7.1	11.5	21.1	30.0	58.7	110.1	123.8
	Undistributed corporate profits (S_u)	2.8	−1.6	3.1	4.4	15.9	16.1	13.0	25.2	14.1	41.0	56.4
	Inventory valuation adjustment	0.5	−2.1	−0.2	−0.6	−5.0	−1.7	0.3	−1.9	−5.1	−12.0	−14.1
	Capital consumption adjustment	−1.3	−0.5	−1.1	−0.1	−4.0	−2.1	−2.3	3.8	1.5	−12.2	−14.7
	Wage accruals less disbursements	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Plus	Government transfer payments to persons (R_{gp})	0.9	1.5	2.7	5.6	14.4	16.2	27.0	37.6	75.9	169.8	184.7
	Net interest paid by government to persons and business (R_{gi})	0.7	1.0	1.2	3.2	4.4	4.6	6.5	7.6	11.2	13.6	16.9
	Interest paid by consumers to business (R_{ci})	1.5	0.5	0.8	0.5	2.3	4.4	7.0	11.1	15.5	22.9	25.0
	Business transfer payments (R_b)	0.6	0.7	0.4	0.5	0.8	1.2	2.0	2.8	4.0	7.0	8.1
Equals	PERSONAL INCOME (PI)	84.9	46.9	77.8	169.8	226.1	308.8	399.7	537.0	801.3	1,253.4	1,382.7
Less	Personal taxes (T_p)	2.6	1.4	2.6	20.8	20.6	35.4	50.4	64.9	115.3	169.0	196.9
Equals	DISPOSABLE PERSONAL INCOME (DPI)	82.3	45.5	75.2	149.0	205.5	273.4	349.4	472.2	685.9	1,084.4	1,185.8
Less	PERSONAL OUTLAYS (PO)	79.1	46.5	72.0	120.4	194.7	258.5	332.3	441.9	635.4	1,004.2	1,119.9
	Personal consumption expenditures (C)	77.3	45.8	71.0	119.5	192.0	253.7	324.9	430.2	618.8	980.4	1,094.0
	Interest paid by consumers to business (R_{ci})	1.5	0.5	0.8	0.5	2.3	4.4	7.0	11.1	15.5	22.9	25.0
	Personal transfer payments to foreigners (R_{pf})	0.3	0.2	0.2	0.5	0.4	0.4	0.4	0.7	1.1	0.9	0.9
Equals	PERSONAL SAVING (S_p)	3.1	−1.0	3.3	28.5	10.8	14.9	17.1	30.3	50.6	80.2	65.9

NOTE: Undistributed profits (S_u) are shown in this table without inventory valuation and capital consumption allowance adjustments. The amount with these adjustments is obtained by adding the sum on the two lines following the line for undistributed profits to the amount given for undistributed profits.

SOURCE: *The National Income and Product Accounts of the United States, 1929–74, Statistical Tables, A Supplement to the Survey of Current Business*, 1977, and *Survey of Current Business*, July 1977.

We may also trace the relationships among the aggregates themselves and see how one is derived from another by adding and subtracting individual items. Thus, as was examined in Chapter 2, we proceed from GNP to net national product, national income, personal income, and disposable personal income through the series of steps shown by the entries at the left of Table 12. The last column of data in this table gives the figures for 1976, figures that also may be derived from Table 11, and the other columns give the figures for other selected years since 1929 to permit the reader to make interyear comparisons.

UNITED STATES ECONOMY—CIRCULAR FLOW AND FUNDAMENTAL IDENTITIES

The same relationships among the aggregates set forth above may also be shown as in Figure 4. Starting with GNP at the far left of the figure and following the central channel through to the far right, we see the various items that are deducted and added in moving from one aggregate to another. More importantly, the circular-flow process depicted here enables us to see the sense in which the total flow of spending on final product, which is equal to

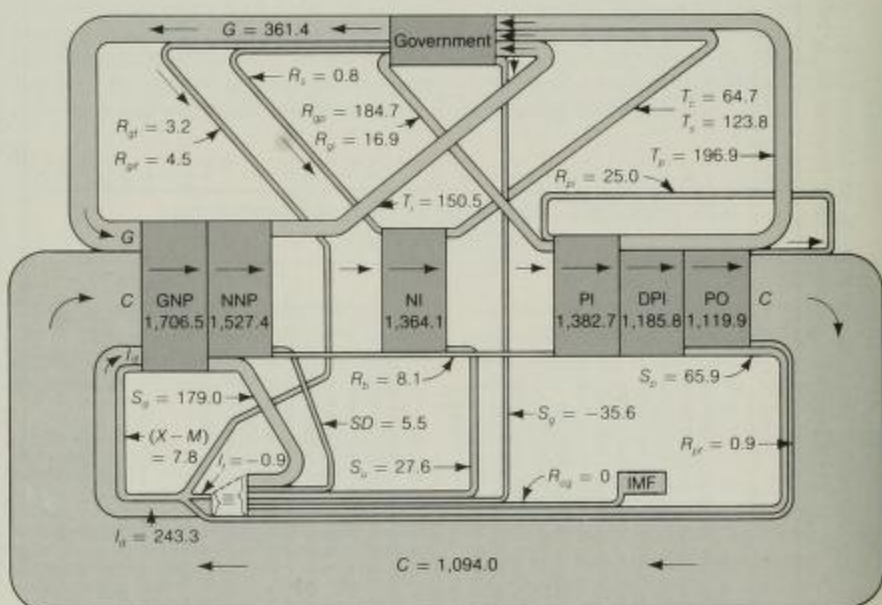


FIGURE 4
Circular flow in the U.S. economy, 1976

GNP, is matched by an equal flow of gross income, as well as the sense in which the portion of this gross income flow that is devoted to saving is matched by an equal portion of gross product devoted to investment.

Using the actual figures (in billions of dollars) for the United States economy in 1976, we have the following for the GNP identity:

$$\begin{aligned} C \quad + R_{pf} + \quad S \quad + \quad T \quad &= \text{GNP} \\ 1{,}094.0 + 0.9 + 278.0 + 333.5 &= 1{,}706.5 \\ &= C + I_g + G + (X - M) \\ &= 1{,}094.0 + 243.3 + 361.4 + 7.8 \end{aligned}$$

In the diagram, GNP of \$1,706.5 on the product side is shown at the far left by total expenditures on final product by the four sectors of demand, $C + I_g + G + (X - M)$. The identity states that corresponding to the \$1,706.5 on the product side, there must be on the income side an identical sum made up of $C + R_{pf} + S + T$. The income flow of \$1,706.5 is split up into $C + R_{pf} + S + T$ as follows. Government extracts from the central flow of income a total of \$535.9 (T_i of \$150.5 plus T_c of \$64.7 plus T_s of \$123.8 plus T_p of \$196.9), but it restores to this central flow a total of \$202.8 (R_s of \$0.8 plus R_{gp} of \$184.7 plus R_{g} of \$16.9). Thus, we can say that government withdraws a *net* amount of \$535.9 − \$202.4, or \$333.5, which equals net taxes, T. Next we must find how much of this gross income flow takes the form of private saving, S. The amount is readily identified by the downward diversions from the central income channel. Here we see that business saves \$206.6 (S_d of \$179.0 plus S_u of \$27.6 in which S_u includes an inventory valuation adjustment of \$−14.1 and a capital consumption adjustment of \$−14.7)[39] and that persons save \$65.9 (S_p), making a total S of \$272.5. We thus have a total of $S + T$ amounting to \$606.0. A final minor diversion from the gross income flow is the one for personal transfer payments to foreigners, R_{pf}, of \$0.9.[40] We then have a total of $S + T + R_{pf}$ of \$272.5 + \$333.5 + \$0.9, or \$606.9. What is left after these diversions is by definition C. Subtracting \$606.9 from \$1,706.5 leaves \$1,099.6, which should be our figure for C in the identity. Looking at Figure 4, however, we see that C is \$1,094.0. This difference is due to the statistical discrepancy (SD) of \$5.5 for 1976, which represents the difference between the Commerce Department's estimates of GNP on the income and product sides of the National Income and Product Account. The discrepancy is treated in effect as an error in the estimate of the diversion into private saving, and the figure for private saving is adjusted accordingly. Therefore, we have \$272.5 + \$5.5, or \$278.0, as the adjusted figure for private saving, S. The total of $S + T + R_{pf}$ thus becomes \$278.0 + \$333.5 + \$0.9, or \$612.4, which, when deducted from the gross income flow of \$1,706.5, leaves the remainder, C, at \$1,094.0, allowing for \$0.1 due to rounding, and the identity is satisfied.

The saving—investment identity, with data taken from Figure 4, is as follows:

$$\begin{aligned} S \; + \quad (T - G - R_{gf}) \quad + R_{cg} \\ 278 + (333.5 - 361.4 - 3.2 - 4.4) + 0.0 \\ = \quad I_d \quad + \quad I_f \\ = 243.3 + (-0.9) \end{aligned}$$

In the GNP identity, neither gross saving (the sum of gross private and public saving) nor gross investment (the sum of gross private domestic and net foreign investment) appears explicitly. The saving—investment identity, however, expresses these factors explicitly in order to show that for every dollar of gross income *not* spent on the purchase of final goods and services by either government or households (i.e., gross saving), there is a dollar of final goods left unconsumed (i.e., gross investment). This is expressed by the identity given above, in which both sides (after adjustment for rounding) are equal to \$242.5.

[39] The inventory valuation adjustment and the capital consumption adjustment will be explained in the following section.

[40] This diversion appears in Figure 4 as the small channel at the bottom of PO. The personal outlay for interest paid by consumers, which appears as the channel at the top of PO, is not a *net* diversion from the central income chan-

V
SOME PROBLEMS IN THE NATIONAL INCOME ACCOUNTS

In the hypothetical economies of Sections I–III, certain problems were ignored in order to simplify our development of the accounting frameworks and the relationships among the product and income flows out of which these frameworks were built. With the real-world economy of the United States, however, it is important to recognize that what is actually measured or, more correctly, estimated by the official income and product accounts depends on the way in which a number of special problems have been resolved by those who prepare the estimates. For many of these problems there is no clear solution, and national income experts differ as to what is the best compromise. In this section we will touch only lightly on a few of the more important of these problems and the treatment given them in the official U.S. accounts.[41]

IMPUTATIONS

From the hypothetical two-sector economy to the actual U.S. economy, we have defined national product as the sum of expenditures for final product and national income as the sum of earnings of the factors of production. All income and product were implicitly assumed to take a monetary form. In the real-world economy, however, restricting income and product to those flows that take explicit monetary form clearly understates the economy's actual income and product. There are nonmonetary income and product flows that would escape inclusion unless certain imputed values were added to the accounts. The nature of these imputations can best be seen by describing a couple of those presently made in the preparation of the official income and product estimates.

Rental Value of Owner-Occupied Homes

A person who rents a home to another earns gross rental income, which becomes net rental income after deduction of expenses.[42] To the tenant, rent paid for the home is an expenditure for the purchase of the services of property. It is part of personal consumption expenditures and therefore part of GNP. For owner-occupied homes, however, there is in monetary terms neither net rental income as part of national income nor rent paid as part of GNP. Since the same real product is provided by a home whether it is owner or tenant occupied, an imputation is made to provide consistent treatment in the accounts. It is assumed for account-

nel, for the part diverted from PO is offset by an equal amount added to PI. This is similar to the amount of taxes diverted to government that is offset by the return flow of government transfer payments to persons, government interest, and subsidies.

[41]For an introduction to other problems of the kind here considered, see S. Rosen, *National Income and Other Social Accounts*, Chapter 8, Holt, Rinehart, and Winston, 1972.

[42]Net rental income of persons is limited to the rental income of persons *not* primarily engaged in the real estate business. Net rental income of those so engaged appears as proprietors' income and, in the case of corporations, as corporate profits.

ing purposes that home ownership is a business in which the owner sells to himself as a tenant the services of the home. The amount of imputed rent is estimated as the sum for which the owner-occupied home could have been rented, and the imputed net rental income is estimated as that portion of the sum that would have accrued to the homeowner after deduction of all expenses. It is thus something like a profit that would be earned by a property owner whose ordinary business is the renting of residential property.

As a result of this imputation, personal consumption expenditure is increased on the product side of the National Income and Product Account by the imputed gross rental for owner-occupied homes. Net rental income of persons, indirect taxes (property taxes), net interest (on mortgages), capital consumption allowances (depreciation of owner-occupied homes), plus other adjustments add an equal amount on the income side of the account. Without this imputation, GNP would vary in response to changes toward or away from home ownership. In any given time period, the larger the fraction of total housing that was tenant occupied, the larger GNP would be. Similarly, GNP would grow larger over successive time periods with a trend away from home ownership toward home rental. By making this imputation, such distortions in GNP are avoided.

Wages and Salaries Paid in Kind

Included under the heading Other Labor Income in Table 11 is an imputed amount for wages and salaries paid in kind. This simply acknowledges the fact that some compensation of employees takes the form of food and lodging provided by employers. No allowance for this labor income would be made in the national accounts if compensation of employees were limited only to monetary income. Accordingly, an estimate is made of wages and salaries paid in kind, and this amount is shown as a supplement to wages and salaries on the income side and as an increase in personal consumption expenditures on the product side. In effect, this accounting treatment assumes that the employer pays his employees the dollar value of the free food and lodging provided and that the employees in turn spend this amount to purchase the same food and lodging from the employer. The imputation thus adds an equal amount to GNP in the form of personal consumption expenditures and to national income in the form of compensation of employees.

Other Imputations

Several other imputations included in the official accounts are those made for food and fuel produced and consumed on farms, and for nonmonetary income and product flows arising from certain operations of commercial banks and financial intermediaries.

There are many other possible imputations that are not included in the official accounts. One of these is production in the form of housewives' services in their own homes. Because there is no statistically sound method of estimating the value of this production, there is no way of including it by imputation, and we must live with the paradox that results from its omission.[43] Thus, the housewife's output, which consists of cooking, washing, cleaning, child care, and other services, is all excluded from the national output. If, however, the housewife were to hire a woman who was previously employed in a factory to do the housework, and she took that woman's place in the factory, GNP would show a net increase equal to the

[43]For an approach to the measurement of this production, see R. Gronau, "The Measurement of Output of the Nonmarket Sector: The Evaluation of Housewives' Time," in M. Moss, editor, *The Measurement of Economic and Social Performance*, Volume Thirty-Eight, Studies in Income and Wealth, National Bureau of Economic Research, 1973. If this production were included, it would appear as compensation of employees on the income side and as personal consumption expenditures on the product side.

amount of the salary paid to the woman hired to perform the housework. Actual output would be unchanged, but the estimate for GNP would be larger. The same paradox applies to work performed by the husband around the house: maintaining the grounds, painting the house, adding a room, or repairing an appliance. None of this production is included in GNP, yet all of it would be included at the full price paid if a landscaper, painter, carpenter, or appliance repairman were hired to do the same work.[44]

With recognition of imputations, both those that are made and those that are not, comes recognition of the frequently overlooked fact that GNP, as officially defined, is an amount that does not purport to be a measure of the economy's output in any all-inclusive sense. By the same token, national income does not purport to be a measure of factor earnings in any all-inclusive sense. Much is excluded from both income and product that would be included under broader definitions.

CHANGE IN INVENTORIES AND THE INVENTORY VALUATION ADJUSTMENT

One part of final product for any year is whatever change in inventories occurs during that year. In 1976 the change was \$13.3 billion, which made final product that much larger than it otherwise would have been for that year. In 1975 the change was −\$11.5 billion, which had a corresponding effect in the other direction.

Figures like \$13.3 or −\$11.5 billion are the Commerce Department's estimates of the physical volume change in inventories for the year, valued at average prices paid for goods added to inventories during the year. This change may be larger or smaller than the book value change in inventories for the year as shown by the accounting data of business firms. The difference between the department's published figure and the figure derived from business accounting data is the *inventory valuation adjustment.* This adjustment, on the product side of the National Income and Product Account, is designed to avoid understating or overstating the change in inventories and therefore to avoid understating or overstating gross domestic investment and gross national product; on the income side the adjustment avoids understating or overstating corporate profits and proprietors' income and therefore national income.

Although the department's methods of estimating the amount of the adjustment are very complicated, the meaning of the adjustment itself is fairly simple. Suppose, for example, that business book values show an increase in inventories of \$3 billion during the year. This figure may be accompanied by an increase, no change, or even a decrease in the physical volume of inventories, since the book value change depends not only on the change in volume but also on changes over the year in the prices paid for inventories acquired. Assume that the change in inventories valued at average prices for the year is only \$1 billion. The discrepancy between the two figures then follows from book values based on prices that are higher than average prices for the year. The Commerce Department would add an inventory valuation adjustment of −\$2 billion to the book value figure of \$3 billion to give a net amount of \$1 billion as the change in inventories valued at average prices for the year.

If prices were falling during the year, the discrepancy between the change in the book

[44]Only that part of production represented by the homeowner's labor services escapes inclusion; the grass seed, paint, lumber, and other materials are included in gross national product in both cases.

value of inventories and the change in inventories valued at average prices would be positive. The change in inventories might show a book value of −$2 billion, but a value at average prices over the year of $1 billion. Accordingly, the inventory valuation adjustment is $3 billion. Regardless of the direction of change in the physical volume of inventories, if prices have risen over the period, the inventory valuation adjustment will necessarily be negative, and if prices have fallen over the period, the adjustment will necessarily be positive.[45]

The inventory valuation adjustment does not appear explicitly on the product side of the National Income and Product Account; instead it is included in the published figure for the change in inventories. The adjustment does appear explicitly on the income side as an adjustment to corporate profits and proprietors' income because it is specifically these two types of income that would otherwise be either overstated or understated. Suppose that the inventory valuation adjustment were −$2 billion, indicating that $2 billion of the change in the book value of inventories was merely the result of valuation at prices higher than the period's average prices. This amounts to an understatement of $2 billion in the amount of noncapital goods used up by business during the period and therefore a $2 billion overstatement of the profits of business. The opposite would, of course, be true of an adjustment of $2 billion.

The inventory valuation adjustment in the official accounts varies from negligible to sizable amounts, depending primarily on the change in the physical volume of inventories and the change in prices at which these inventories are valued by business. For example, as a percentage of national income, it was an extraordinarily large 3.7 in 1974, 0.7 in 1970, and a negligible 0.01 in 1961. However, whether large or small in any particular year, it is an item explicitly shown in the official framework, and some understanding of its nature is helpful to an understanding of the complete official framework.

CAPITAL CONSUMPTION ALLOWANCES AND THE CAPITAL CONSUMPTION ADJUSTMENT

As shown by Item 25 in Account I of Table 11, the difference between charges against GNP and charges against NNP or, what is the same thing, the difference between GNP and NNP is capital consumption allowances with capital consumption adjustment. Prior to the introduction of the capital consumption adjustment in the 1976 revisions of the national income and product accounts, the difference between the two aggregates was capital consumption allow-

[45] For any time period, the size of the inventory valuation adjustment varies with the particular methods used by business in charging inventories against cost of goods sold. Under one method [First In, First Out (FIFO)], inventories are charged against cost of goods sold in the order that inventories were acquired. Under another method [Last In, First Out (LIFO)], inventories are charged against cost of goods sold in reverse of the order that inventories were acquired. If both prices and the volume of inventories are rising, FIFO shows an inventory increase that reflects both the increase in volume and the higher prices paid for the last inventories acquired. For, with inventories carried in the books at original cost, FIFO includes the lower-cost inventories acquired earlier as a cost of goods sold during the period and includes the higher-cost inventories acquired later in the book value of inventories at the end of the period. The change in book values from the first to the end of the period thereby overstates the increase in volume, and a negative inventory valuation adjustment is required. If prices are falling and the volume of inventory is rising, the change in book values under FIFO understates the change in volume, and a positive inventory valuation adjustment is required. Whether prices are rising or falling, as long as volume is rising, LIFO results in neither overstatement nor understatement of the change in volume and no inventory valuation adjustment is required. However, under LIFO, if the volume is decreasing, the decrease is then being measured not in prices paid for those inventories acquired last but in prices paid for those inventories acquired earlier, and an adjustment is required.

ances only. These allowances consist of depreciation charges and accidental damage to fixed business capital, and the amount of these allowances is determined primarily from the depreciation charges reported in the federal income tax returns of firms. As the figure for capital consumption allowances is therefore essentially the amount of depreciation recorded in the books of business, it serves as an acceptable measure of the amount of the business capital stock actually used up to the degree that there is a correspondence between the bookkeeping entries for depreciation and the actual wear and tear on capital.

There are a number of reasons why these amounts may not closely correspond. For one, depreciation reported in tax returns is based on asset lives and depreciation formulas that may not agree very closely with the using up of plant and equipment. For example, Internal Revenue Service regulations may understate the true service life of various kinds of assets and may permit accelerated depreciation formulas that compress much of the total depreciation into the first few years of the lives of those assets. The changes in regulations during the postwar period appear to have resulted in write-offs of assets at a more rapid rate than assets have been used up. A second shortcoming is that the depreciation recorded in tax returns is based on the original or historical costs of assets, and it therefore values different assets on different price bases, depending on the various years in which assets were acquired. A figure for the value of capital goods consumed by all firms in 1976 should not be influenced by the particular mixture of prices that is given by the happenstance of the years in which different assets were acquired, but that is just the result that follows from the use of depreciation charges.

To avoid the shortcomings requires an alternative that is not based on tax-return depreciation charges. In 1976 the Department of Commerce incorporated into the national income and product accounts new estimates based on an alternative method. These estimates are derived from stocks of fixed capital calculated by the perpetual inventory method. This method uses estimates of gross investment and service lives of capital goods to derive estimates of the gross stock of capital goods. The gross stock in any year is thus found as the cumulative gross investment of prior years less the gross investment in fixed assets that have completed their service lives. Capital consumption allowances are derived for any one year as a portion of the investment elements that remain in the gross stock. This amount is obtained by applying a straight-line depreciation formula to the stocks of various classes of fixed assets with the life for each class being its service life. The valuation basis for assets is not the original or historical cost but replacement cost.[46]

The new measure of capital consumption allowances is greater than the old in some years and less than the old in other years. *The new measure less the old equals a new item called the capital consumption adjustment.* For example, in 1976 the new measure was \$179.0 billion and the old was \$142.0 billion so that the capital consumption adjustment was \$37.1 billion, the \$0.1 discrepancy being due to rounding. What the adjustment figure says is that the old or tax-return based measure of \$142.0 billion understated the amount of capital used up by \$37.1 billion.

The capital consumption adjustment, whatever its size, has no effect on GNP or gross private domestic investment. However, it changes the amount subtracted from GNP to arrive at NNP and the amount subtracted from gross private domestic investment to arrive at net private domestic investment, and it therefore does affect NNP and net private domestic investment. In Account I of Table 11, the capital consumption adjustment of \$37.1 billion makes the figure for Item 25 that much larger than it would otherwise be and thus makes NNP

[46]For a full explanation of this method, see A. H. Young, "New Estimates of Capital Consumption Allowances in the Benchmark Revision of GNP," *Survey of Current Business*, October 1975.

that much smaller than it otherwise would be. Under the old measure, NNP would have been $1,564.5 billion instead of the $1,527.4 billion shown. The introduction of the capital consumption adjustment also changes the estimates of business income. As shown in Account I, proprietors' income, rental income of persons, and corporate profits each includes a capital consumption adjustment. The adjustment appears explicitly only for corporate profits: $14.7 billion of the 1976 adjustment of $37.1 was in this share. Part of the balance, $3.8 billion, was in proprietors' income and another part, $16.7 billion, was in rental income of persons. In the case of corporate profits, the $14.7 billion adjustment means that the $128.1 billion figure given in Item 10 is $14.7 billion lower than it would have been under the old tax-return based measurement basis. The same kind of conclusion may be drawn for the other two income shares noted.

It will be seen that the treatment here is similar to that described above for the inventory valuation adjustment. The inventory valuation adjustment is necessary to correct for an overstatement or understatement of profit-type incomes that results from the way that firms value their inventories. The capital consumption adjustment also corrects for an overstatement or understatement of profit-type incomes but for one that results from the failure of tax-return depreciation charges to measure the actual wear and tear on the capital stock.

The new measure of capital consumption avoids the shortcomings inherent in the old tax-return based measure and is regarded by the Department of Commerce as a measure of "economic" capital consumption and thus one that permits the calculation of meaningful figures for NNP and net fixed private domestic investment. As mentioned in Chapter 2, heretofore the figures for GNP and gross private domestic investment have been the ones generally used, but this may change in the years ahead with the availability of the new economic measure of capital consumption.

FINAL PRODUCT—CURRENT AND CONSTANT DOLLARS

Final product is equal to the sum of expenditures for those goods and services that are defined as final goods and services. The total represents the dollar value of the output of final goods and services produced by the nation's economy, with each unit of final product valued at the actual market price paid for it.

Using this procedure in measuring final product, an increase or a decrease in the dollar value of final product from one year to the next will result from changes either in market prices paid or in the volume of goods and services purchased, or in both. The usual purpose of comparing figures for final product over time is to determine changes in the volume of goods and services and, with qualifications, changes in the economic welfare of the economy. Yet, if the final product figures for each year are expressed in prices current in each year, such comparisons are practically ruled out. Plainly, we are faced with the task of eliminating from these figures the effects of price changes so as to reveal the actual volume changes over time.

One apparently easy method of accomplishing this might be to divide any year's final product, valued in that year's prices, by a price index such as the Consumer Price Index. By doing so one expresses the value of that year's final product as the dollar amount it would have been if prices had been the same as those in the base year of the index. For example, suppose that the Consumer Price Index, equal to 100 in the selected base year 1967, was 120 in the year 1971. This would indicate that con-

sumer prices were 20 percent higher in 1971 than in 1967. Suppose further that 1967 final product valued in 1967 prices is \$200 and 1971 final product valued in 1971 prices is \$300. Since we know that prices have risen during these years, this increase of \$100 in final product is clearly not altogether due to an increase in the volume of product. How do we determine what part of the increase is attributable to higher prices and what part to higher volume? One way of solving the problem is to divide 1971 final product valued in 1971 prices by the ratio of 1971 prices to 1967 prices.

$$\frac{\$300}{120/100} = \$300 \times \frac{100}{120} = \$250$$

This tells us that of the increase in final product of \$100 measured in 1971 prices, half was due to an increase in volume and half to a rise in prices from 1967 to 1971. In other words, final product in 1971 valued in 1967 prices was \$250. Final product for other years could be converted from prices current in each of those years to the prices in effect in the base year of 1967 in the same fashion, so that final product for each year would be expressed in "constant dollars of 1967 purchasing power," here measuring changes in the purchasing power of the dollar by the Consumer Price Index.[47]

In our numerical example, we may reliably conclude that the volume of final goods and services in 1971 was not 50 percent greater than that in 1967 (\$300/\$200 = 1.50). Can we also reliably conclude that it was 25 percent greater than that in 1967 (\$250/\$200 = 1.25), as indicated by the 1971 dollar value of \$300 corrected to \$250 by the Consumer Price Index? The answer is clearly no, for all that has been done is to adjust the 1971 market value of all kinds of diverse goods and services, including new factory buildings, aircraft carriers, machine tools, salaries of government employees, net exports, and everything else that is part of final product, by a single price index. This price index is at best appropriate only to that part of final product made up of the goods and services purchased by consumers. For the same reason, adjustment of 1971 final product by the Wholesale Price Index or any other single price index would be unreliable.

The fact that no single price index is appropriate suggests that one must break down final product as finely as possible and then adjust each part by a price index appropriate to the goods and services included in it. This is essentially the procedure followed by the Commerce Department.[48] Thus, personal consumption expenditures are broken down into dozens of parts, and for each part a special price index is prepared if an acceptable one is not already available. A similar procedure is followed for government purchases of goods and services, gross domestic investment, and exports and imports. In each case the dollar value of each

[47]Designating final product by GNP, it follows that in any given year (g), GNP is that year's volume of output (O_g) multiplied by the prices paid for each unit of that output (P_g), or $GNP_g = O_g \times P_g$. Expressing prices of any given year as a percentage of prices in the base year gives us a price index in which base-year prices are equal to 100. In other words, the index number for any given year is P_g/P_b, where b designates the base year. Finally, the GNP figure for any given year may be converted into a figure for the output of the given year valued in base-year prices by dividing it by the price index for the given year, as shown by the equation

$$\frac{GNP_g}{P_g/P_b} = \frac{O_g \times P_g}{P_g/P_b} = O_g \times P_g\frac{P_b}{P_g} = O_g \times P_b$$

Performing this operation for each given year yields a series of final figures for GNP in which output, O_g, is valued in base-year prices, P_b. Comparison of these adjusted GNP figures will presumably show only changes in volume, the effect of changes in prices having been eliminated.

[48]Where there are ways of measuring directly the change in volume of a final good or service over time (e.g., in gross domestic investment, the number of freight cars of a given description purchased by railroads, or in government purchases of goods and services, the man-hours of labor, or the number of persons of a given class employed), the physical quantity purchased in any year multiplied by the base-year price per unit will equal the value of that final good or service for the year in base-year prices. This technique bypasses the need for a price index. If data on volume are available, the department uses this method whenever it appears that by doing so it can obtain more accurate results.

TABLE 13
Gross national product in current and constant (1972) dollars, selected years, 1929–76 (billions of dollars)

(1) Year	(2) Current Dollars	(3) Constant Dollars	(4) Implicit Price Deflator (1972 = 100)
1929	103.4	314.7	32.87
1933	55.8	222.1	25.13
1937	90.7	309.8	29.29
1939	90.8	319.7	28.40
1945	212.3	559.0	37.99
1948	259.1	487.7	53.13
1950	286.2	533.5	53.64
1954	366.3	613.7	59.69
1957	442.8	680.9	65.02
1958	448.9	679.5	66.06
1962	563.8	799.1	70.55
1965	688.1	925.9	74.32
1966	753.0	981.0	76.76
1967	796.3	1,007.7	79.02
1968	868.5	1,051.8	82.57
1969	935.5	1,078.8	86.72
1970	982.4	1,075.3	91.36
1971	1,063.4	1,107.5	96.02
1972	1,171.1	1,171.1	100.00
1973	1,306.6	1,235.0	105.80
1974	1,412.9	1,217.8	116.02
1975	1,528.8	1,202.1	127.18
1976	1,706.5	1,274.7	133.87

SOURCE: *The National Income and Product Accounts of the United States: 1929–74, Statistical Tables, a Supplement to the Survey of Current Business*, 1977, and *Survey of Current Business*, July 1977.

part for the given year is divided by its price index for that year; the resulting dollar values are added together, and the sum is taken as the given year's GNP valued in base-year prices.

This procedure is called "deflation" of GNP. The deflated GNP is termed GNP in "constant dollars" to distinguish it from undeflated GNP, which is in "current dollars." The use of the term "deflation" may be misleading, for to deflate GNP in current dollars to get GNP in constant dollars suggests that the constant dollar figure must be less than the current dollar figure. This, however, is true only in those years in which the prices of the goods and services comprising GNP are higher than those prices in the base year. For other years it is necessary, in effect, to "inflate" the current dollar figures to obtain the constant dollar figures. Nonetheless, by convention, the term deflation is applied in both cases.

The base year prices used are those of 1972. Since prices in the years preceding 1972 were lower than those in 1972, the constant dollar figure for each such year is greater than the current dollar figure; since prices in the years following 1972 were higher than those in 1972, the constant dollar figure for each such year is less than the current dollar figure. The relationship between constant-dollar GNP and current-dollar GNP may be seen in Table 13, which gives GNP in current and constant dollars for selected years from 1929 to 1976.[49]

As a by-product of the department's deflation procedure, there emerge what are officially designated "implicit price deflators" for GNP. These are given in Column (4) of Table 13. For example, the deflator for 1973, 105.80 is "implicit" in the current and constant dollar figures for 1973 in that it is derived by dividing 1,306.6 by 1235.0 and multiplying the quotient by 100. Column (4) turns out to be a price index but one quite different from the Consumer or Wholesale Price Index. Unlike the more specific coverage of other price indexes, the implicit price deflator covers all the varied goods and services included in the economy's GNP.

The department's deflation procedure yields more reliable results than those secured by any of the short-cut methods described earlier. Nevertheless, even though the official constant-dollar estimates are probably as accurate as present data and techniques can make them,

[49]Although GNP in constant dollars remains the most widely used measure of real aggregate output and until 1976 was the only such measure available, NNP and NI as well as other aggregates are now available in constant as well as current dollars. Data for these series as well as the implicit price deflators for each will be found in the same sources noted for Table 12.

some unknown error is unavoidably present in each estimate. In large part, this error results from the shortcomings of the price indexes themselves—for example, the problem of allowing in indexes for quality changes in the goods included and of adjusting for the appearance of new goods and the disappearance of old. Whatever error may be present, these estimates still provide the most dependable indicator currently available for the actual changes over time in the physical volume of goods and services included in GNP.

VI NATIONAL PRODUCT AND NATIONAL WELFARE

Although those who produce the official estimates of GNP and the other measures of income and output have long told the users of these figures that they are not designed as measures of national welfare, to many of these users this very fact constitutes a major shortcoming of the national income accounts. In other words, to them the failure of an aggregate like GNP to measure national welfare indicates a fault in the way the GNP is estimated. As they see it, after adjustment for changes in prices and population, a 5-percent increase in the figure for GNP should reflect a 5-percent increase in national welfare and vice versa, something which it surely does not do as presently constituted.

No one criterion of national welfare is universally accepted, but the one that is most widely adopted is that which holds that the national welfare is enhanced by an expansion of goods and services designed to satisfy the needs of ultimate consumers today and in the future.[50] Although we did not look at them from this particular point of view, we have seen in the last few pages various illustrations of why national welfare expressed in terms of this criterion and national output as defined by the Department of Commerce will diverge. That definition now excludes from GNP some goods and services that contribute to an enhancement of national welfare and includes others that do not.

What is now excluded from GNP that should be included if changes in GNP are to reflect changes in national welfare? The major omissions are those services that do not involve market transactions and that are also not picked up by imputations, the foremost case of which is the value of housewives' services. These services surely qualify for inclusion under the above criterion, but the difficulty is, as we saw earlier, that there is no statistically sound method of putting a dollar and cents value on them. Of course, each time a housewife cuts back her specific role as housewife to take a position in the world of business, we do have a value placed on at least part of her services and we thus do include the market value of these services in GNP. This means a greater figure for GNP than would otherwise be the case, and, on the basis of the most reasonable approach to the question, the increase in GNP is accompanied by an increase in welfare. If we accept the argument that the individual is the best judge of what increases or decreases his or her welfare, it follows that the very fact that a housewife voluntarily takes a position in business means that to her that brings with it an increase in welfare. If the other members of her family

[50] When the criterion includes the needs of consumers in the future as well as the present, the goods that qualify to meet the needs are capital goods as well as consumer goods and services. Capital goods increase the economy's capacity to produce more consumer goods and services for consumers in the future.

also find an increase (or at least no decrease) in their welfare as a result of this action, her outside employment involves an unambiguous increase in family welfare.

But even though this increase in welfare will be reflected in a change in GNP that is in the right direction, the change in GNP will apparently seriously overstate the increase in welfare. No dollar value at all was placed on the housewife's services in the home, whereas her services to business were counted at the full market value paid for them. Some subtraction from this latter figure is required to allow for the services no longer performed or neglected in the home. There is, however, no way of adjusting for this without placing a dollar and cents value on the housewife's services in the home, and this, of course, is the stumbling block that prevents the inclusion of these services in GNP in the first place. Although GNP is caused in this way to overstate the increase in welfare that presumably results from housewives' taking outside employment, the far more serious problem presented by this same stumbling block remains the one discussed earlier. The complete exclusion from GNP of all of the services performed within the home means that GNP is, as a result of this exclusion alone, a very poor measure of welfare, if we take as our criterion that welfare is enhanced by an expansion of goods and services that satisfy the needs of consumers. The amount of these services provided within homes grows over time more or less in line with the growth in the number of families, but none of this is reflected in the changes in GNP from year to year.

What is now included in GNP that should be excluded if changes in GNP are to reflect changes in national welfare? Of the many areas in which this problem appears, the major one is that of government purchases. At present, GNP includes all government purchases of goods and services, whatever their nature. From the standpoint of national welfare, many of these purchases should be excluded because they do not meet the welfare criterion accepted above—they do not add to the output of goods and services designed to satisfy the needs of consumers today and in the future. Most notable is government expenditure for national defense. In the view of many people, an expansion in this area of spending means a contraction of national welfare through the diversion of resources that could otherwise be used to produce goods and services that satisfy the needs of consumers. However, under present definitions, each dollar so spent appears as an additional dollar of GNP, other things being equal. As these people view the result, national welfare and national product seem to move in opposite directions in this case.

Although the unusually strong opposition to military spending during the late sixties and early seventies was in large part attributable to the unfortunate Vietnam experience, in other times there have been arguments that such spending should be included as part of GNP. Simon Kuznets, one of the leading students of national income measurement, argued during the years before World War II for the exclusion of government military outlays (as well as certain other classes of government purchases) because they do not meet the criterion noted earlier. However, during World War II he modified the criterion and reversed his position on the grounds that national survival becomes an end of economic activity that should be included along with the individualistic goal of satisfying consumer wants. In the years since the war he has probably gone back to his original criterion with regard to military expenditures. However, other reasonable men may view the large military budgets of the postwar years (excluding the especially controversial Vietnam costs) as a matter of national survival in the world as it exists and thus argue that the portion of GNP accounted for by such expenditures is matched by a corresponding contribution to national welfare.

It would be possible to go down the list of government purchases, military and all others, and somehow classify them into those that contribute to consumer satisfaction, present or future, and those that do not and then include in

GNP only those that do. Although the GNP would thereby presumably come closer to being a measure of national welfare, it would be a measure far more influenced by personal judgments and biases than the present measure. Which government expenditures should be included or excluded might vary from time to time as conditions changed (as in the World War II illustration) and as the decision-makers themselves changed. Instead of getting a measure whose changes reflect in any meaningful way changes in national welfare, we might get a measure whose changes measure something that nobody can any longer identify at all accurately. Little would be gained, and the reasonably definite measure we now have would be lost.

But even if these problems that appear to be unsolvable were solved and the national income accountants somehow were able to identify and include in GNP all of those goods and services that are designed to satisfy the needs of consumers, present and future, and identify and exclude from GNP all those that are not, the GNP figure so constituted would still not be a significant indicator of national welfare. It would be so only to the extent that national welfare can be measured by the amount of consumer-satisfying goods and services produced. There are numerous and serious shortcomings to this total as a measure of welfare.

An increase in the total of such goods would ordinarily indicate an increase in national welfare, but how much of an increase would depend very much, for one thing, on the distribution of the GNP or, as this is usually expressed, on the distribution of income. A more equal distribution of the existing level of income and output might make a greater contribution to national welfare than a sizable increase in the total with no change in its distribution. The approach to improving welfare via income redistribution as compared to income expansion is one that has received increasing attention in recent years.

Even if there were no change in the distribution of income, we still could not conclude that welfare is reflected at all accurately by the output of goods. A major consideration is not only the amount of goods produced but also the "human cost" of producing that amount. A given amount will mean greater or lesser welfare depending on the human cost of its production. Mechanization of arduous and repetitive tasks, improvements in work safety, and revised production techniques that restore "pride of workmanship" are examples of things that reduce the human cost. But what is probably the largest single factor under this heading is the number of hours of work per week put in by the average worker. If a given total of output can be secured with a substantially smaller number of hours of labor input, it clearly would be incorrect to say that the constancy of output means a constancy of welfare. To the unchanged output of goods must now be added the additional "output" in the form of additional leisure. Leisure, like goods, satisfies the needs of consumers, but unlike goods, leisure is not counted in the GNP. Like housewives' services in the home, leisure has value, but in this case too there is no reasonably determinate way of putting a dollar and cents figure on it. The amount of this kind of "output" has greatly increased over the years, and its exclusion from the GNP is one of the major reasons that GNP cannot serve as a measure of welfare.

If a dollar value could be placed on a change in the distribution of income as it is on the output of goods, if a dollar value could be placed on the output of leisure as it is on the output of goods, and if a myriad of other similar problems could be solved, the resultant GNP figure would have moved a long way toward becoming a measure of welfare. However, it would still be such a measure only in terms of a criterion that makes national welfare a matter of the amount of output produced and its distribution or essentially a matter of materialism. And this is, as everyone will agree, an extremely narrow

criterion of national welfare. There are any number of other things quite unrelated to the output of goods and services that could make the nation better off: peace, improved race relations, equal educational opportunity for all, elimination of crime and violence, a greater voice for minorities, an improved system of courts and justice, and the like.

The possibility of ever being able to put a reasonable dollar value on the things discussed earlier is small enough; the possibility of doing this for hundreds of things like those just noted is far less. Yet this is what would have to be done to convert GNP into a meaningful measure of national welfare. And even then we would not really be in the clear. We will not pursue this aspect, but according to one school of thought, which includes economists like John Kenneth Galbraith and E. J. Mishan, the criterion set forth above—that national welfare is enhanced by an expansion of goods and services designed to satisfy the needs of consumers, present and future—is itself unacceptable, so the whole structure built on it has no real foundation. The basic argument of this school is that the developed countries of the world have already reached a state at which all rational needs for goods and services have been or can be satisfied and that mere expansion of goods and services does not add to national welfare. The argument is also extended to hold that we have reached a situation in which the further expansion of goods may involve a decrease rather than an increase in welfare. The deterioration in environmental quality and the reduction in amenities that result from producing more and more output are costs that exceed the benefits of having that greater output. When we charge the negative amounts for ecological damage against the positive amounts represented by the goods themselves, the balance is negative, or a decline in welfare with growth in goods. This conclusion leads some to support zero economic growth, ZEG, or a nonexpanding level of national output hereafter. There is much to be said for the arguments of this school, and the critics have much to say against them. Some of these arguments were touched on in the discussion of the relationship between economic growth and environmental quality in Chapter 20.

The part of this that is directly relevant to the question before us is the following: If one reaches the point at which he must question whether even an expansion of national output of goods and services increases or decreases national welfare, there can be little doubt remaining that a measure that does no more than put a dollar value on the national output of goods and services is a far cry from a measure of national welfare. And this is essentially all that the measure called gross national product now does and, in view of all the obstacles considered above, all that it will likely do in the future. There is clearly a great need to develop measures or indexes of national welfare, if this can be done, but it is unlikely that GNP or any other such aggregate will evolve into such a measure. It will probably continue to be what it is now and to do what it does now: Provide a reasonably good measure of the value of the output of final goods and services, where final goods and services are defined in the way examined in some detail in this appendix.

Index

O

P

Q

R

S